The impact of sex on gene expression across human tissues

Oliva, Meritxell; Muñoz-Aguirre, Manuel; Kim-Hellmuth, Sarah; Wucher, Valentin; Gewirtz, Ariel D. H.; Cotter, Daniel J.

Published in:
Science (New York, N.Y.)

DOI:
10.1126/science.aba3066

Publication date:
2020

Document Version
Peer reviewed version

Title: The impact of sex on gene expression and its genetic regulation across human tissues

Authors: Meritxell Oliva[1,2,3*†], Manuel Muñoz-Aguirre[4,5†], Sarah Kim-Hellmuth[6,7,8†], Valentin Wucher[4], Ariel D.H. Gewirtz[9], Daniel J. Cotter[10], Princy Parsana[11], Silva Kasela[7,8], Brunilda Balliu[12], Ana Viñuela[13], Stephane E. Castel[7,8], Pejman Mohammadi[14], François Aguet[15], Yuxin Zou[16], Ekaterina A. Khramtsova[1,17], Andrew D. Skol[1,2,18,19], Diego Garrido-Martín[4], Ferran Reverter[20], Andrew Brown[21], Patrick Evans[22], Eric R. Gamazon[22,23], Anthony Payne[24], Rodrigo Bonazzola[1], Alvaro N. Barbeira[1], Andrew R. Hamel[15,25], Angel Martinez-Perez[26], José Manuel Soria[26], GTEx Consortium, Brandon L. Pierce[3], Matthew Stephens[16,27], Eleazar Eskin[28], Emmanouil T. Dermitzakis[13], Ayellet V. Segrè[15,25], Hae Kyung Im[1], Barbara E. Engelhart[9], Kristin G. Ardlie[15], Stephen B. Montgomery[10,29], Alexis J. Battle[11,30], Tuuli Lappalainen[7,8], Roderic Guigó[4,31], and Barbara E. Stranger[1,2,18, 32*]

† Contributed equally to this work

* Corresponding author. Email: meritxellop@uchicago.edu, barbara.stranger@northwestern.edu

Affiliations:

[1] Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL, USA

[2] Institute for Genomics and Systems Biology, The University of Chicago, Chicago, IL, USA

[3] Department of Public Health Sciences, The University of Chicago, Chicago, IL, USA

[4] Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Barcelona, Catalonia, Spain

[5] Department of Statistics and Operations Research, Universitat Politècnica de Catalunya (UPC), Barcelona, Catalonia, Spain

[6] Statistical Genetics, Max Planck Institute of Psychiatry, Munich, Germany

[7] New York Genome Center, New York, NY, USA

[8] Department of Systems Biology, Columbia University, New York, NY, USA

[9] Department of Computer Science, Center for Statistics and Machine Learning, Princeton University, Princeton, NJ, USA

[10] Department of Genetics, Stanford University, Stanford, CA, USA

[11] Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA

[12] Department of Computational Medicine, University of California, Los Angeles, CA, USA

[13] Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland

[14] Department of Integrative Structural and Computational Biology, The Scripps Research Institute, Scripps Research Translational Institute, La Jolla, CA, USA

[15] The Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA.

[16] Department of Statistics, The University of Chicago, Chicago, IL, USA

[17] Computational Sciences, Janssen Pharmaceuticals, Spring House, PA, USA

[18] Center for Translational Data Science, The University of Chicago, Chicago, IL, USA

[19] Department of Pathology and Laboratory Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA

[20] Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona. Barcelona, Spain

[21] University of Dundee, Dundee, Scotland, UK

[22] Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

[23] Clare Hall, University of Cambridge, Cambridge, England, UK

[24] Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK.

[25] Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA

[26] Genomics of Complex Diseases Group. Research Institute Hospital de la Sant Creu i Sant Pau. IIB Sant Pau. Barcelona, Spain.

[27] Department of Human Genetics, The University of Chicago, Chicago, IL, USA

[28] Departments of Computational Medicine, Computer Science, and Human Genetics, University of California, Los Angeles, CA, USA

[29] Department of Pathology, Stanford University, Stanford, CA, USA

[30] Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA

[31] Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain

[32] Center for Genetic Medicine, Department of Pharmacology, Northwestern University, Chicago, IL USA

Abstract: Many complex human phenotypes exhibit sex-differentiated characteristics, however the underlying molecular mechanisms of these differences remain largely unknown. Here, we present an extensive catalog of both sex differences in gene expression and its genetic regulation across 44 human tissue sources surveyed by GTEx (v8 release). We demonstrate that sex strongly influences gene expression levels and cellular composition of tissue samples across the human body. The effect of sex on gene expression is widespread, with a total of 37% of all genes exhibiting sex-biased expression in at least one tissue. This suggests that many if not most biological processes, and thus complex traits and diseases, are impacted by sex effects on the transciptome. We expand the identification of *cis*-eQTLs with sex-differentiated effects and characterize their cellular origin. By integrating sex-biased eQTLs with genome-wide association study data, we identify 58 gene-trait associations that are driven by genetic regulation in a single sex, including novel associations not detected with sex-agnostic approaches. Altogether we provide the most comprehensive characterization of sex differences in the human transcriptome and its regulation to date.

One Sentence Summary: Sex differences in the human transcriptome are widespread, highly tissue-specific, and contribute to our understanding of sex-differentiated biology and complex traits.

Main Text:

Results

Many complex human phenotypes exhibit sex-differentiated characteristics, including anthropometric traits and disease features such as prevalence, progression, age of onset, and response to treatment (*1*–*5*). Although these sex differences have been previously attributed to hormones, sex chromosomes, and various genetic models (*6*), the mechanisms and contributing biology remain largely unknown, in part due to limited tools and datasets to conduct analyses and interpret sex-related results. The Genotype-Tissue Expression (GTEx) project provides an opportunity to investigate the prevalence and genetic mechanisms of transcriptomic sex differences and how sex and genetics interact to influence complex traits and disease.

Here, we present the largest characterization to date of sex differences in the human transcriptome across 44 tissue sources of the GTEx project (v8 data release) from 838 individuals (Fig. 1A). We have quantified and characterized sex-biased gene expression and *cis* sex-biased expression quantitative trait loci (sb-eQTLs) (Fig. 1B). By incorporating the results of these sex-aware analyses of GTEx data with gene and genome annotation, we describe tissue-specific and tissue-non-specific drivers and mechanisms contributing to sex differences in the human transcriptome. Finally, by integrating data from genome-wide association studies (GWAS), we report multiple sex-differentiated genetic effects on the transcriptome that colocalize with complex trait associations.

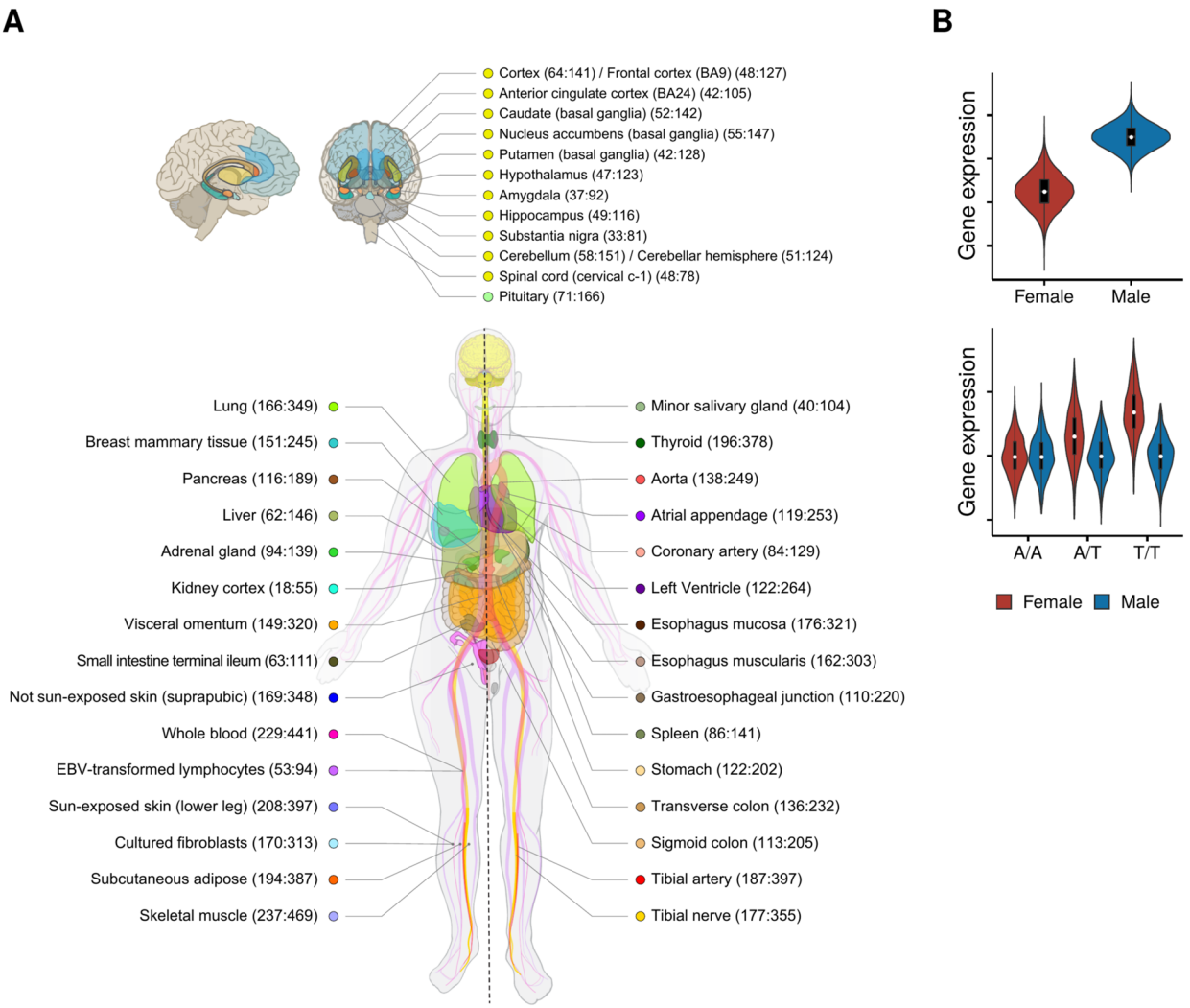


Fig. 1. Sample, data types and discovery sets in the study of sex differences in GTEx v8. (**A**) Illustration of the tissue types (including 11 distinct brain regions and 2 cell lines), with sample numbers from GTEx v8 genotyped donors in parentheses (females:males) and color coding indicated in the adjacent circles. N=44 tissue sources present in both sexes with ≥70 samples were included in this study. Tissue sources comprise two cell lines and 40 different tissues; with two additional replicates for brain cerebellum and cortex tissues. (**B**) Schematic representation of a sex-differentially expressed gene (sex-biased gene) and a sex-biased eQTL (sb-eQTL).

Sex-differentiated gene expression

Using GTEx v8 data (Table S1), we quantified sex-biased gene expression in each of the 44 tissue sources for all genes expressed in at least one tissue comprising a total of 35,422 X-linked and autosomal genes, including protein-coding, lincRNA, and other less characterized gene types. For each tissue, we first fit a linear model using *voom-limma* (*7*) that accounted for known sample and

donor characteristics, as well as surrogate variables that capture hidden technical or biological factors of expression variability, including tissue cell type composition (fig. S1, A and B). We then applied *MASH (8)* to model across-tissue covariation of gene expression and discovered a total of 13,227 differentially expressed genes (sex-biased genes) (local false sign rate (LFSR) ≤ 0.05), with 365 to 4,070 genes discovered per tissue (Fig. 2A, fig. S1, C to E, Table S2). Previous studies reported breast as the most sex-differentiated tissue (*9*), however this is not observed in the present study due to the analysis approach that accounts for sex-differences in tissue cell type composition (fig. S1A). In total, 37% of the human transcriptome was differentially expressed in at least one tissue, of which 545 (4%) were X-linked and 12,682 were autosomal genes. X-linked genes with higher expression in females (female-biased genes) exhibited larger sex effects (median fold change (FC) = 1.20) than either X-linked genes with higher expression in males (male-biased genes) (median FC = 1.11), or autosomal sex-biased genes (median FC_M and FC_F = 1.06) (fig. S2, A and D) potentially due to incomplete dosage compensation or X-inactivation. Overall, the number of sex-biased genes and the effect size was not dominated by either sex (fig. S2B).

Tissue-sharing of sex-biased genes was highly skewed towards sharing across few tissues, with 3,158 sex-biased genes (23.8%) discovered in only a single tissue (Fig. 2B), suggesting tissue-dependent regulation. Only 32 genes (0.24%; 22 of which are known constitutive X-inactivation escapees, Table S3) exhibited consistent sex bias across all 44 tissue sources (Fig. 2B). This tissue-specificity did not simply reflect patterns of gene expression across tissues; sex-biased genes tended to be ubiquituously expressed across tissues while the sex bias was limited to one or few tissues (Fig. 2C, fig. S2C). Although hierarchical clustering based on gene expression or sex-differential expression produced highly concordant tissue clustering (Fig. 2C, fig. S3, A and B), the cluster-defining gene sets were mostly disjoint (Table S4). For example, both data types supported a cluster of brain subregions that is clearly differentiated from other tissues (Fig. 2C, fig. S3, A and C), however the cluster based on sex-biased expression was driven by 60 genes, while the transcriptome-based brain cluster was driven by 631 genes, none of which were drivers of the sex-bias-based brain cluster. Notably, whole blood and cell lines were not representative of sex-biased expression across tissues (fig. S3A), and whole blood represented only 11.7% (1,548/13,277) of all sex-biased genes.

We accurately predicted sex from gene expression using X-linked genes (fig. S4, A to C), as previously described (*10*). Although the most predictive X-linked genes (fig. S4D) are X-inactivation escapees (*11*), we identified 57 highly predictive (top tertile) X-linked genes, not previously described as escapees (Table S3). Sex prediction based on autosomal genes was less accurate, less specific (mean specificity = 56%) (fig. S4C), and required more genes (fig. S4E) than prediction based on X-linked genes; however, in two tissues, breast and muscle, autosomal genes predicted sex with ≥ 90% specificity (fig. S4F).

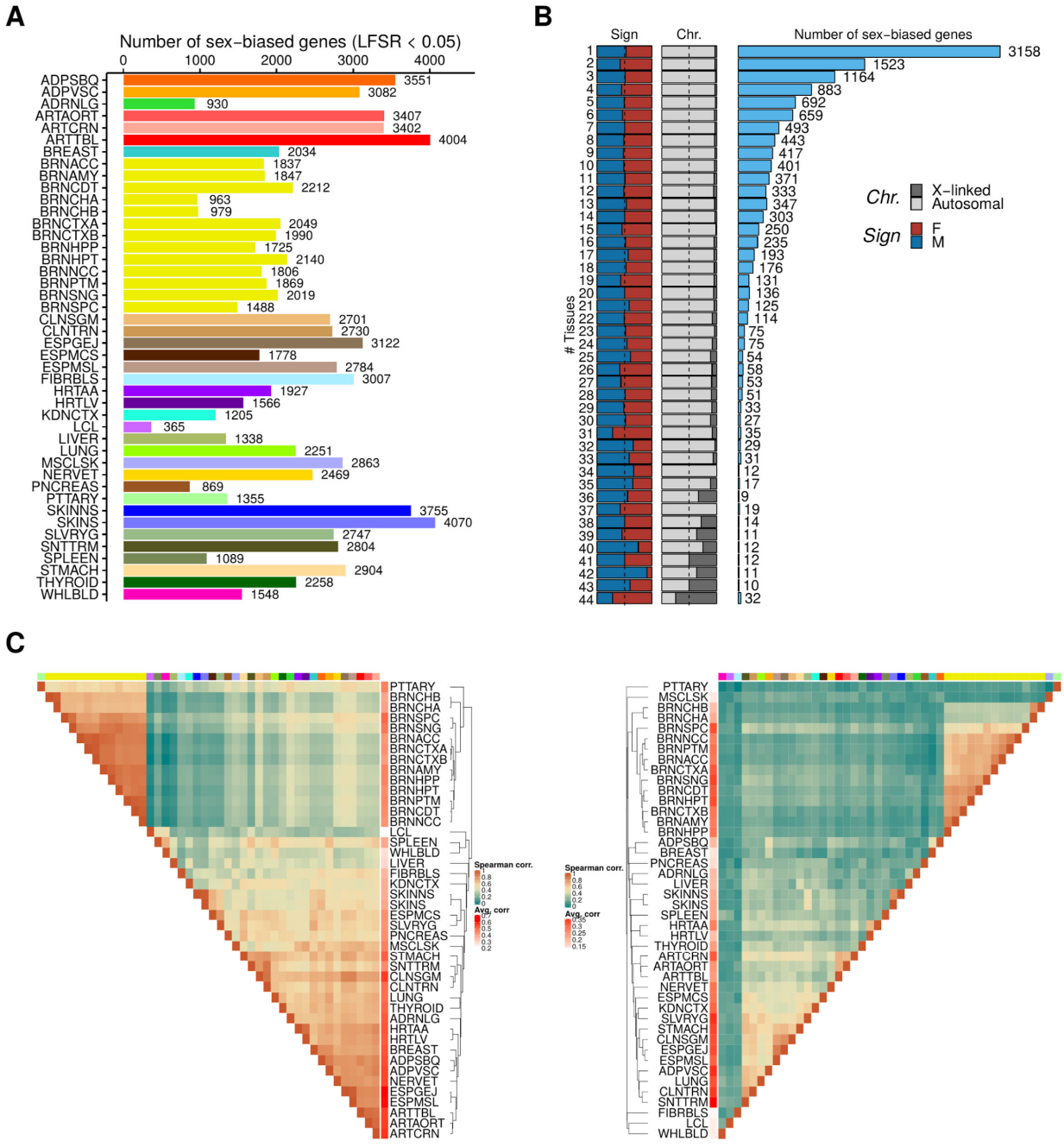


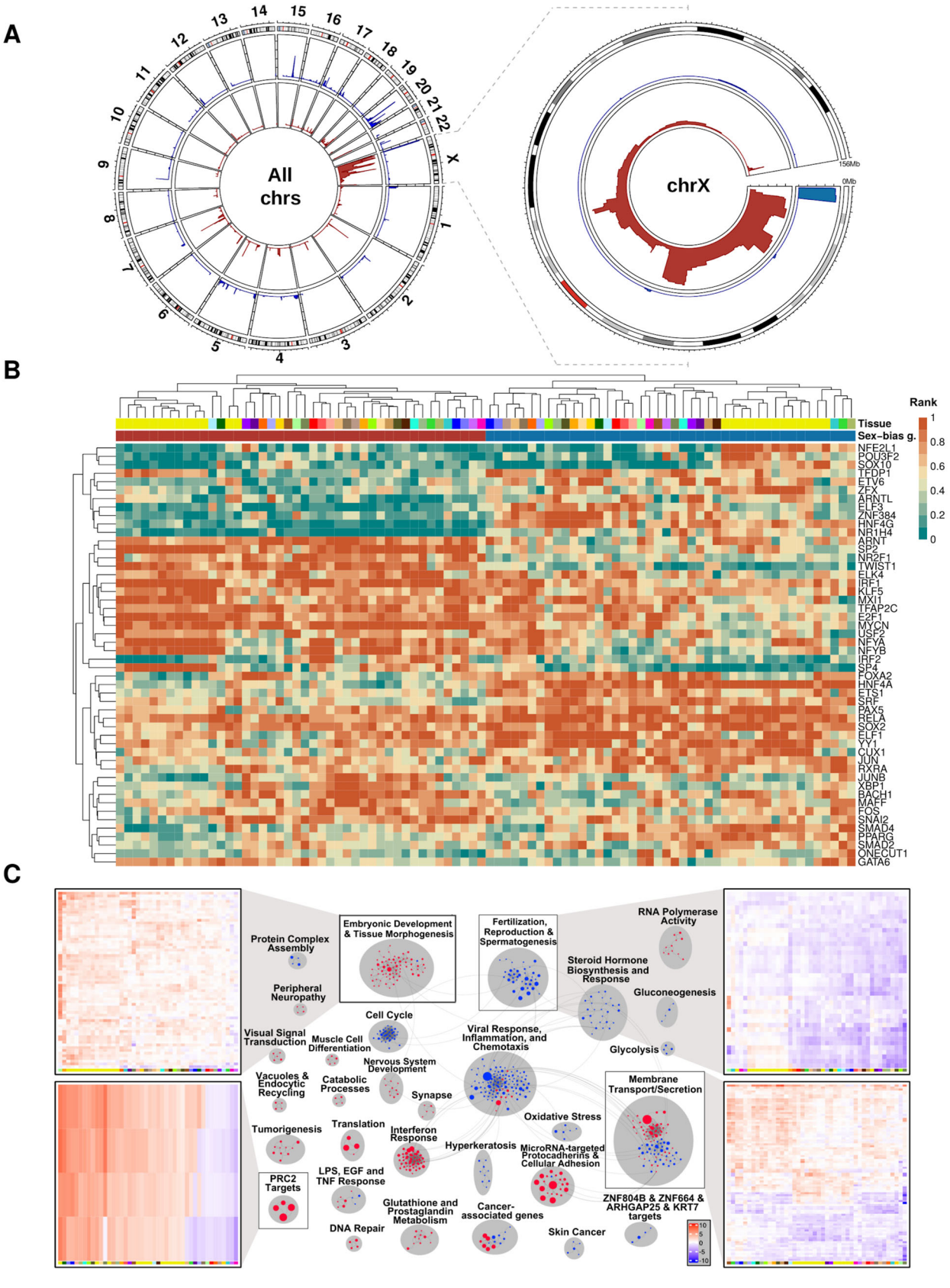
Fig. 2. Sex-differential gene expression. (**A**) Number of sex-differentially expressed genes (sex-biased genes) per tissue. See Fig. 1A for the legend of tissue colors. (**B**) Sex-biased gene discovery (histogram, number of sex-biased genes) and characteristics of sex-biased genes (stacked bar plots) as a function of tissue-sharing. The characteristics are: *Chr.* for the proportion of X-linked and autosomal sex-biased genes, *Sign* for the proportion of female-biased (F) and male-biased (M) genes. (**C**) Hierarchical clustering of tissues based on gene expression (left) and the effect size of sex-biased genes (right).

Except for the enrichment of female-biased genes on the X chromosome, little is known about the genome-wide distribution of sex-biased genes. We applied a positional gene enrichment analysis method (PGE) (*12*) separately for male- and female-biased genes (LFSR ≤ 0.05) from each tissue. We discovered significant clustering of 7,557 sex-biased genes in 159 autosomal and two X-linked regions (hypergeometric test P ≤ 0.001; Fig. 3A, left; Table S5). On ChrX, two regions in the pseudoautosomal (PAR) regions PAR1 and PAR2, were enriched for sex-biased genes (Fig. 3A, right). The strongest enrichments were observed for non-PAR regions, particularly the ChrX short arm *p*, driven by escape from X-inactivation (*11*). Although the region spanned ~134Mb, only 30% of subregions were enriched in at least two thirds of the tissues, confirming tissue-variable escape from X-inactivation (*13*). We observed a cluster of male-biased genes on chromosome 20 that was shared in 75% of tissues (fig. S5A), but the majority of the 159 autosomal enriched regions were strongly tissue-specific with average sharing in ~10% of tissues (fig. S5, B and C, Table S5).

We hypothesized that transcription factor (TF) activity might drive differential expression. We tested for enrichment of TF binding sites (TFBS) of 231 TFs previously identified through ChIP-seq (*14*) in promoter regions (2kb upstream of the transcription start site) of male- and female-biased genes. We discovered enrichment for TFBS of a total of 101 TFs, three of which were X-linked (*AR, GATA1, ZFX*) (Table S6). TFBS for 55 TFs were enriched among female-biased genes and 73 TFs among male-biased genes; with 27 TFs enriched among both (fig. S5D), including known hormone-related TFs estrogen (*ESR1*), androgen and glucocorticoid (*N3C1*) receptors, and also other TFs with a non-reported or less characterized hormone association, such as heat-shock factors (*HSF1*) and *CTFC* (fig. S5D). *CTCF,* a gene with an established role in DNA looping and gene repression, modulates estrogen receptor (ER) function (*15*) and links sex-biased distal regulatory elements to sex-biased expression (*16*), suggesting a possible mechanism for the sex-bias observed here. The strongest sex-bias difference comparing male- versus female-biased enrichment profiles is observed for TFBS of *MYCN, SP2, IRF1, TWIST1, E2F1* (top five female-biased) and *HNF4A, RELA, SOX2, PAX5, POU5F1* (top five male-biased) TFs (Fig. 3B, Table S6), which were detected broadly across most tissues. In contrast, we observed marked tissue-specificity for several TFs, particularly in brain, e.g. *SP4, HIF1A* and *NFE2L1, SMAD3* for female- and male-biased genes, respectively (Fig. 3B, fig. S5D). Together, these results suggest that hormone-related and X-linked TFs regulate sex-biased expression as expected, but also suggest a

role for other, not previously reported, TFs (Table S6). The tissue-specific TFBS enrichments reflect tissue-specificity of differential expression and suggest tissue-specific regulators of sex differential expression.

To gain insight into the functions affected by sex-biased genes, we performed gene set enrichment analysis (GSEA) (Table S7), separately for male-biased and female-biased genes in each tissue (Table S8). To identify gene sets that are enriched across multiple tissues, we performed a meta-analysis using Fisher's combined probability test and identified 1,274 significant (FDR ≤ 0.05) gene sets (Table S9). We applied a community detection approach to identify common features across enriched gene sets and defined 30 clusters (Table S9). Among the top significant (FDR ≤ 1e-05) clusters, we identified pathways involved in cancer, immune response, steroid biosynthesis and response, prostaglandin metabolism, and other functions (Fig. 3C, Table S9). We observed a female-biased cluster of 65 gene sets enriched broadly across tissues and related to embryonic development and tissue morphogenesis (Fig. 3C, top left), and a cluster of 39 male-biased gene sets related to fertilization, sexual reproduction and spermatogenesis (Fig. 3C, top right). Within clusters, gene sets with opposite sex bias were common, e.g., the membrane transport/secretion cluster was comprised of both male-biased and female-biased gene sets (Fig. 3C, bottom right). We identified clusters with notable tissue-specificity, e.g., a cluster composed of four *PRC2* and H3K27-trimethylation target gene sets with strong female-biased or male-biased expression depending on the tissue (Fig. 3C, bottom left). *PRC2* plays an essential role in X-inactivation (*17*). H3K27-trimethylation is an epigenetic mark of gene silencing, and has recently been shown to be deposited in a sex-specific manner by *PRC2* in mouse liver (*18*), and hypothesized to be largely regulated by the sex differences in secretion of growth hormone by the pituitary gland, in concordance with pituitary enrichment patterns observed here (Table S8). Pituitary specific sex-biased profile (Fig. 2C, fig. S3A) was partially explained by hormone signal; 8% (4/49) of genes driving pituitary-specific clustering patterns (Table S4) are part of hormone-related gene sets, and among those, genes *NTS*, *QRFPR* and *PTGFR* are related to prostaglandin pathways (Table S8). These analyses revealed that sex-biased genes are enriched in many biological functions and pathways, including functions both known and unknown to exhibit sex differences.


A
All chrs
chrX
156Mb
0Mb
B
Rank
1
0.8
0.6
0.4
0.2
0
Tissue
Sex-bias g.
NFE2L1
POU3F2
SOX10
TFDP1
ETV6
ZFX
ARNTL
ELF3
ZNF384
HNF4G
NR1H4
ARNT
SP2
NR2F1
TWIST1
ELK4
IRF1
KLF5
MXI1
TFAP2C
E2F1
MYCN
USF2
NFYA
NFYB
IRF2
SP4
FOXA2
HNF4A
ETS1
SRF
PAX5
RELA
SOX2
ELF1
YY1
CUX1
JUN
RXRA
JUNB
XBP1
BACH1
MAFF
FOS
SNAI2
SMAD4
PPARG
SMAD2
ONECUT1
GATA6
C
Protein Complex Assembly
Embryonic Development & Tissue Morphogenesis
Fertilization, Reproduction & Spermatogenesis
RNA Polymerase Activity
Steroid Hormone Biosynthesis and Response
Peripheral Neuropathy
Gluconeogenesis
Cell Cycle
Visual Signal Transduction
Muscle Cell Differentiation
Viral Response, Inflammation, and Chemotaxis
Glycolysis
Nervous System Development
Vacuoles & Endocytic Recycling
Catabolic Processes
Synapse
Membrane Transport/Secretion
Oxidative Stress
Tumorigenesis
Translation
Interferon Response
Hyperkeratosis
MicroRNA-targeted Protocadherins & Cellular Adhesion
PRC2 Targets
LPS, EGF and TNF Response
ZNF804B & ZNF664 & ARHGAP25 & KRT7 targets
Glutathione and Prostaglandin Metabolism
Cancer-associated genes
DNA Repair
Skin Cancer
10
5
0
-5
-10

1

Fig. 3. Regulatory mechanisms and biological functions of sex-biased genes. (**A**) Genomic position enrichment of sex-biased genes; based on male-biased (blue), and female-biased (red) genes across all chromosomes (left) and chromosome X (right). The y-axis represents the tissue-sharing of the significant genomic enrichment signal and ranges from 1 to 44 (number of tissue sources)**.** See Methods for further details. (**B**) Transcription Factor Binding Site (TFBS) enrichment in promoter regions of sex-biased genes, for those TFBSs with the strongest difference (top 50) in the enrichment profile derived from male-biased and female-biased genes across all tissues. Values represent the TFBS enrichment ranking transformed to [0, 1]. (**C**) Clusters of gene sets (balloons) enriched for genes highly expressed in females (red) or males (blue) across tissues. Balloon size corresponds to the p-value for the across-tissue meta-analysis of GSEA. For framed clusters, heatmap values correspond to signed GSEA p-values of gene set (rows) per tissue (columns) pairs, enriched for genes highly expressed in females (red) or males (blue).

Sex differences in tissue cell type composition

The GTEx tissue samples are comprised of mixtures of heterogeneous cell types, with variation among individuals and tissues (*19*). In whole blood, cell type composition differs between sexes (*20*, *21*), but little is known about sex differences in composition of other tissues. We tested each GTEx tissue for sex differences in cellular composition using estimated abundances of seven cell types (*19*). We discovered significant (FDR ≤ 0.05) differences for four cell types in three tissues (Fig. 4A, Table S10). In skin exposed to sun, keratinocytes were more abundant in males (P = 4.46e-03). In whole blood, neutrophils were more abundant in females (P = 1.52e-06). In breast, males exhibited a higher abundance of adipocytes (P = 1.84e-08), while epithelial cells were markedly more abundant (P = 8.46e-24) in females, where ducts, stroma, and glandular tissue predominate. To investigate cellular abundances in disease, we utilized histological annotations based on pathology review of GTEx tissue samples. We discovered six pathological phenotypes with altered cell type composition (Fig. 4B, fig. S6, Table S10). Male breast samples diagnosed with gynecomastia (N = 69), a male-specific condition involving an enlargement or swelling of breast tissue, exhibited more epithelial cells and fewer adipocytes than non-diagnosed samples ($P_{\text{Epithelial cells}}$ = 1.91e-14, $P_{\text{Adipocytes}}$ = 1.71e-14, Fig. 4C), as previously reported (*22*). Gynecomastia is characterized by altered levels of steroid hormones and their receptors (*23*), as well as TRH-mediated prolactin levels (*24*)*;* among other genes, *TRH* expression in breast differed between males with and without gynecomastia (Fig. 4D; P = 2.20e-14). We also discovered altered cell type composition for other conditions, including a lower abundance (P = 1.90e-03) of epithelial cells in female thyroid samples affected with Hashimoto's thyroiditis (N = 25), an autoimmune disease more prevalent in females (*4*) that is characterized by destruction of thyroid epithelial cells (*25*), in agreement with our observations. Adipocytes were more abundant in liver samples

annotated with steatosis (N_M = 65, N_F = 31), as expected (*26*), however, this difference was significant only in males (P_M = 3.99e-05, P_F = 0.22), although association in females was in the same direction (β_M = 0.63, β_F = 0.30).

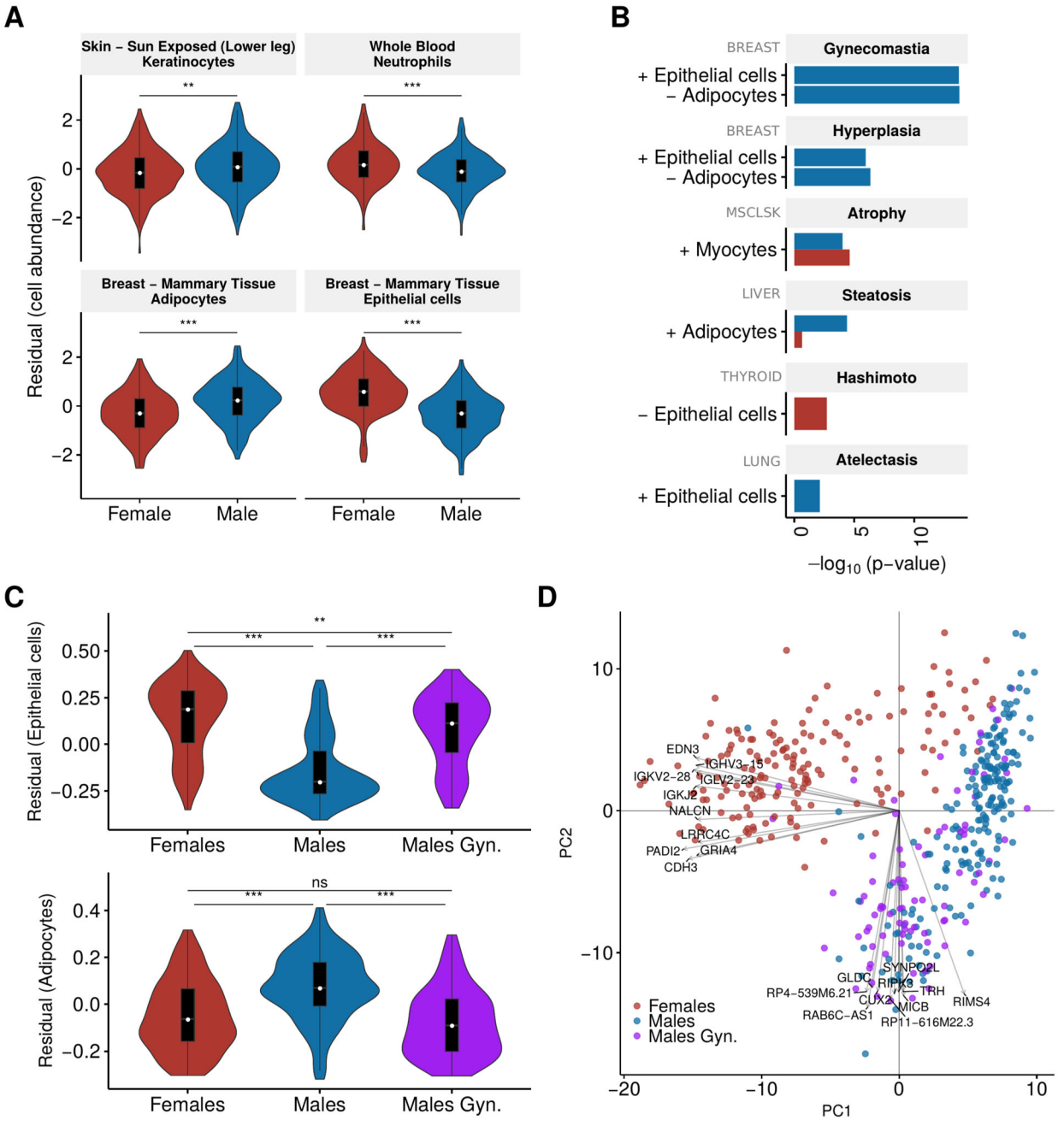


Fig 4. Sex differences in cellular abundances. (**A**) Sex differences in estimated cell type abundances (y-axis) within tissues. Values correspond to normalized and residualized xCell enrichment scores from (*19*). (**B**) Barplots depicting significant associations of histological phenotypes with cell type abundances by tissue; x-axis corresponds to the -

log10(p-value) of the association corrected for multiple testing (Methods). Direction of association is indicated by the sign, where a positive sign represents higher abundance in pathological samples and a negative sign, the opposite. (**C**) Distribution of estimated cell type abundances (y-axis) in breast samples of females, males, and males with gynecomastia. The significance levels used for (A) and (C) are (*): ≤ 0.05, (**): ≤ 0.01, (***): ≤ 0.001. (**D**) Principal component analysis (PCA) of breast transcriptome using sex-predictive autosomal genes. Samples are color-coded by sex and gynecomastia status. The top ten most correlated genes with each principal component (PC1, PC2) are labeled.

Sex-differences in genetic regulation of gene expression

Sex-differentiated human phenotypes and disease characteristics may derive, in part, from sex-differentiated genetic effects (*6*, *27–30*), some of which may impact gene expression. For 491,694 conditionally independent *cis*-eQTLs identified in combined male and female *cis*-eQTL analysis in the GTEx v8 project (*31*), we performed sex-biased *cis*-eQTL (sb-eQTL) analysis in each of 44 tissues present in both sexes (Fig. 1A). We employed a linear regression model testing for significance of genotype-by-sex (G×Sex) interaction on expression, while adjusting for known cofactors and unknown covariates; modeled with *FastQTL* (*32*). We discovered a total of 369 sb-eQTLs, corresponding to 366 genes (sb-eGenes) (FDR ≤ 0.25, Table S11, fig. S7, A and B). The majority of sb-eQTLs were identified in breast (261 sb-eQTLs), but also in muscle (36 sb-eQTLs), skin (18 sb-eQTLs) and adipose tissues (14 sb-eQTLs) (Fig. 5A). Only 53 (14%) sb-eGenes exhibited significant sex-biased expression in the discovery tissue (*MASH* LFSR ≤ 0.05) (Table S12). To provide additional support for the sb-eQTLs, we used two approaches to assess differential allele-specific expression (ASE) between sexes: allelic fold-change (ASE aFC) (*33*) and EAGLE (*34*). Allele-specific expression can arise due to *cis*-regulatory genetic effects in heterozygous individuals, and differential ASE, therefore, supports condition-specific *cis* effects (*34*) such as sex-specificity. We observed that, despite limited power when restricted to heterozygous individuals and differences in methodology, both approaches support that a portion of the detected sb-eQTLs correspond to sex differences in ASE (fig. S7C): sb-eQTLs were enriched for sex-biased ASE aFC (All tissues: $\pi_1 = 0.36$, breast: $\pi_1 = 0.41$, fig. S7, D and E) and for EAGLE associations ($\pi_1 = 0.13$, empirical test, P ≤ 0.001). Of the 243 and 163 sb-eQTLs tested by ASE aFC and EAGLE methods, respectively, 65 (26.7%) were supported by significant ASE aFC (Wilcoxon P ≤ 0.05) (fig. S7, F and G), 29 (17.8%) were supported by significant EAGLE associations, and 16 (10.4% of 154 sb-eQTLs tested by both methods) were supported by both (Table S11).

We are limited in our ability to replicate sb-eQTLs because the majority of sb-eQTLs were discovered in breast, and matching, well-powered datasets do not exist. We performed internal validation, splitting GTEx breast samples into discovery and validation cohorts, and observed moderate replication (mean π_1 = 0.28) (fig. S7H). We also assessed sb-eQTL replication (considering sb-eQTLs from breast, whole blood, and all tissues) in independent whole blood eQTL datasets, including DGN (*35–37*) and GAIT2 (*38*) cohorts (Table S13), and observed weak replication (π_1 = 0-0.12, depending on sb-eQTL set and replication cohort). Low replication of sb-eQTLs has been previously reported (*35–37*) and has been attributed, in part, to low power (*39*). For each sb-eGene, we also performed sex-stratified *cis*-eQTL analysis for each tissue, downsampling males to match the female sample size, and observed strong correlation (ρ = 0.78, P ≤ 2.2e-16) between male and female *cis*-eQTL effect sizes. For 58% of sb-eQTLs, sex-stratified *cis*-eQTL analysis revealed significant associations in both sexes with concordant allelic effect, but different effect sizes, e.g., rs117380715-*ADRA1A* in adipose subcutaneous tissue (β_F = -0.78, β_M = -0.47) (Fig. 5B). For the remainder, a *cis*-eQTL was detected exclusively in either females (70, 19%) or males (84, 23%). For example, we identified a female-specific *cis*-eQTL for rs8942-*C4BPB* in breast (β_F = 0.40, β_M = -0.02) (Fig. 5B). *C4BPB* encodes the beta unit of C4b-binding protein, and controls activation of the complement cascade. We also identified a male-specific *cis*-eQTL for rs2273535-*AURKA* in skeletal muscle (β_M = 0.47, β_F = 0.01), described in (*31*). *AURKA*, Aurora kinase A, is a member of the serine/threonine kinase family involved in mitotic chromosomal segregation, and a known risk factor for several cancers (*40*), and also involved in muscle differentiation (*41*). These results demonstrate that sex-biased genetic effects on gene expression exist for a small proportion of previously identified *cis*-eQTLs, and some impact sex-differentiated phenotypes. Overall, sb-eQTLs are strongly tissue-specific; only one sb-eQTL was significant in two tissues (Table S11), and only 21% displayed patterns suggestive of tissue-sharing even at a lenient significance threshold (nominal P ≤ 0.01, not shown).

We hypothesized that a portion of the sex-differentiated genetic effects on gene expression may derive from sex differences in cell type composition in the tissue (fig. S8A). To test this, we focused on breast, the tissue with the most sb-eQTLs and sex differences in cellular composition (Fig. 4A). We tested 261 breast sb-eQTLs for enrichment of cell type interacting *cis*-eQTL (ieQTL) signal (*19*). These ieQTLs correspond to *cis*-eQTLs whose effect varies depending on the

cell type abundance estimates (*19*). Breast sb-eQTLs were strongly enriched (π_1 = 0.66 and 0.89) for ieQTL signal corresponding to both adipocytes and epithelial cells (fig. S8B). After including an interaction term for genotype-by-epithelial cell abundances in the sb-eQTL model, 58% of breast sb-eQTLs (152/261) remained significant, while for 42% of sb-eQTLs (109/261), the genotype-by-sex effect was strongly attenuated (fig. S8C, Table S14). For example, the strongest breast sb-eQTL, rs2289149-*LINC00920* (P = 4.83E-11) was not significant after incorporating the genotype-by-epithelial cell abundances in the model (β = 0.187, CI(95%) = [-0.004 , 0.378], fig. S8C, Table S14). To formally test the impact of cell type composition on sb-eQTL detection, we performed a mediation analysis considering genotype interactions with epithelial cell abundances as a potential mediator (fig. S8D), and discovered that 60 (23%) sb-eQTLs were significantly mediated (ACME P $\leq$ 0.001, Fig. 5C) (Table S14). Mediation by other cell types beyond epithelial cells cannot be excluded, and for the significantly mediated sb-eQTLs, our analysis cannot distinguish whether the tested cell types or others correlated with them (fig. S8E) are the true mediators of the signal. Nevertheless, these results show that a large proportion of sb-eQTLs are driven by a combination of cell type-specific genetic regulatory effects and cell type differences between sexes.

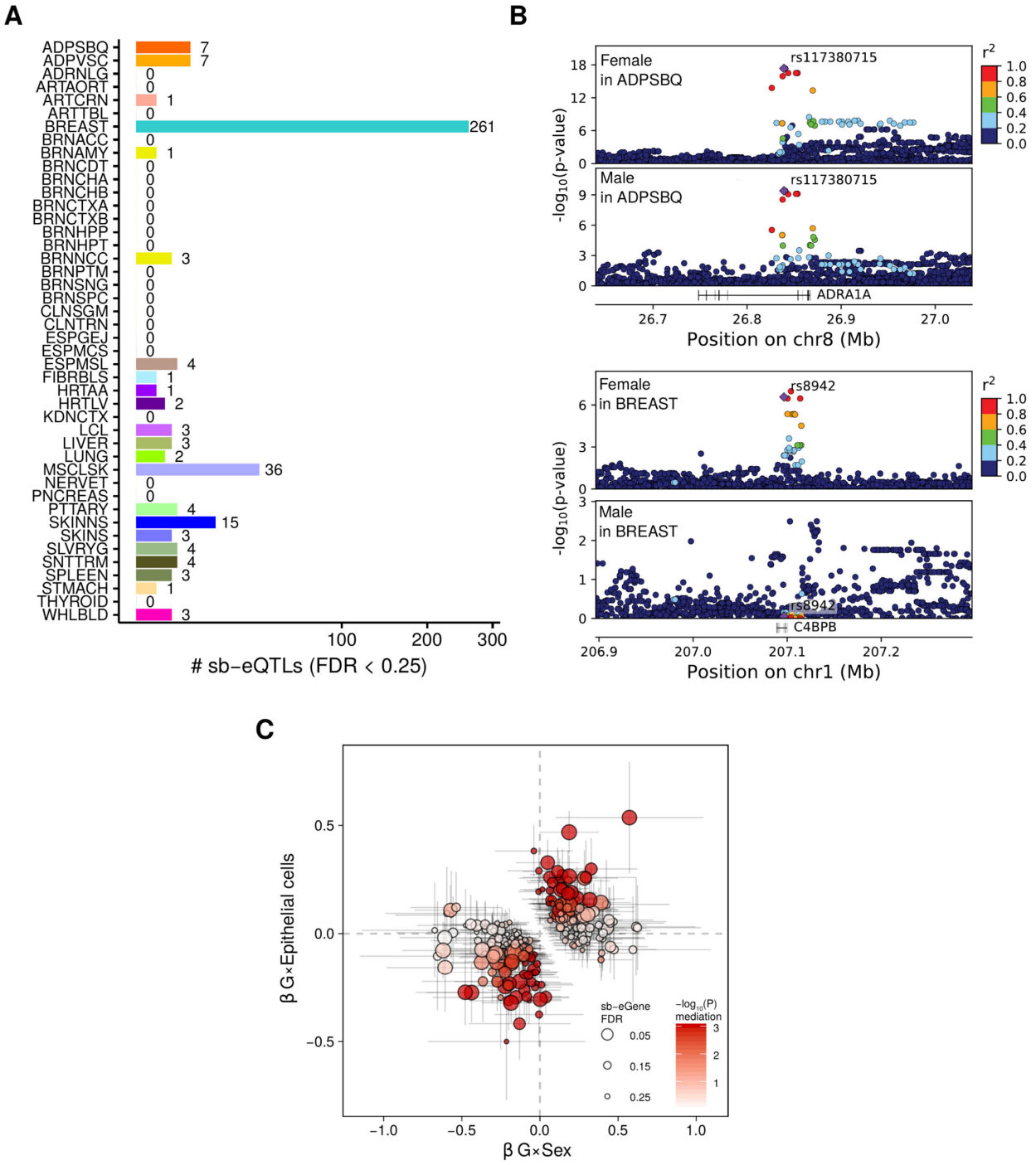


Fig. 5. Sex-biased eQTLs (sb-eQTLs). (**A**) Number of sb-eQTLs discovered per tissue. Square-root transformation was applied to the x-axis. (**B**) P-value landscapes of the female- (top) and male-stratified (bottom) *cis*-eQTLs in the *ADRA1A* locus (top panels, β_F = -0.78, P_F = 4.64e-18, β_M = -0.47, P_M = 3.98e-10, $P_{G\times Sex}$ = 1.05e-05) and *C4BPB* locus (lower panels, β_F = 0.40, P_F = 2.68e-07, β_M = -0.02, P_M = 8.90e-01, $P_{G\times Sex}$ = 7.22e-05). (**C**) sb-eQTL mediation analysis of 261 breast sb-eQTLs. Points represent the effect sizes of the terms G×Sex (x-axis) and G×Epithelial cells (y-axis) derived from a linear regression model with both interaction terms. Point size represents sb-eQTL significance and color corresponds to mediation significance.

To assess the utility of sb-eQTLs to dissect the basis of complex traits, we performed colocalization (*42*) between sex-stratified *cis*-eQTLs and complex traits for 1,089 sb-eGenes at a more relaxed FDR (FDR ≤ 0.50). We identified 74 colocalized gene-trait pairs (posterior probability of sharing the same causal variant, PP4 > 0.5) (Fig. 6A), of which 58 were colocalized (PP4 > 0.5) in one sex but not in the other - 36 for females and 22 for males - that correspond to 42 unique loci (Fig. 6, A and B, Table S15). For 24/36 (67%) female- and 10/22 (45%) male-stratified *cis*-eQTL-trait pairs, colocalization was also found using the male and female combined GTEx v8 *cis*-eQTLs (fig. S9A). For these 34 loci, we provide evidence that the colocalization signal is driven by regulatory effects in a single sex. For the remaining 12/36 (33%) female and 12/22 (55%) male gene-trait pairs, the GWAS colocalizations are novel discoveries. The strongest colocalization between traits and female-stratified *cis*-eQTLs was identified for *CCDC88C* and *HKDC1* genes (Fig. 6C). Conversely, the strongest colocalization between traits and male-stratified *cis*-eQTLs was identified for *DPYSL4* and *CLDN7* genes (Fig. 6D). *CCDC88C* is a negative regulator of the Wnt signalling pathway, a key mechanism in cancer progression (*43*) and the *CCDC88C* female *cis*-eQTL signal in breast colocalizes with risk of breast cancer (Fig. 6C, left). For breast cancer, we identified two additional female-driven colocalized sb-eGenes, *NTN4* and *CRLF3* (Table S15), previously reported as breast cancer-relevant genes (*44*, *45*). Additionally, we discovered a significant preferential colocalization of blood and immune traits with female- compared to male-driven *cis*-eQTLs (odds ratio = 2.22, Fisher's exact test P = 0.047); including inflammatory bowel diseases, which show a higher prevalence in females with increasing age (*46*) and immune cell abundances in blood, which also exhibit sex differences (*20*, *21*). Together, these results suggest that sex-biased regulation of gene expression may contribute to the etiology of diseases with marked sex differences in prevalence. Moreover, we identified colocalization signal for sex-specific traits. For instance, the *HKDC1* female *cis*-eQTL signal in liver colocalized with birth weight, which is strongly influenced by maternal factors (Fig. 6C, right). The sb-eQTL for this locus in liver was replicated in an external cohort (*47*) (rs35696875-*HKDC1* P_F = 2.73e-08, P_M = 1.60e-04, z-test P = 0.004, fig. S9B). *HKDC1* encodes a member of the hexokinase protein family and is involved in glucose metabolism. Multiple variants in perfect or high LD with rs35696875 that cause reduced expression of *HKDC1* have been associated with glycemic traits during pregnancy (*48*) and gestational diabetes mellitus risk (*49*). Recently, it has been shown that regulatory variants spanning multiple enhancers have a coordinated allelic effect on *HKDC1*

expression (*50*). These results suggest that the *HKDC1* female *cis*-eQTL influences glucose metabolism in the pregnant mother, which is reflected in the birth weight of the child. The *DPYSL4* male *cis*-eQTL signal in skeletal muscle colocalized with percentage of body fat (Fig. 6D, right). *DPYSL4* is linked to the pathophysiology of obesity and cancer: p53-inducible *DPYSL4* associates with mitochondrial supercomplexes and regulates energy metabolism in adipocytes and cancer cells and low *DPYSL4* expression is significantly associated with poor survival of breast cancer patients (*51*). Of note, while the colocalizing signal was detected with the female-only *cis*-eQTL signal, the low probability of colocalization appears to be due to the presence of an additional *cis*-eQTL in females that was absent in males. These results suggest that characterizing sex differences in the genetic associations of high-order traits and molecular phenotypes can prove useful to dissect allelic heterogeneity. For five colocalized sb-eGenes (*CLDN7, CCDC125, FAM53B, PLEC, SOWAHC*), corresponding cell type interaction *cis*-eQTL (cell type ieQTL) signal also colocalized with reported GWAS traits (birth weight, blood cell counts, height, platelet counts and schizophrenia, respectively) (*19*). For instance, the male-biased *cis*-eQTL rs34958987-*CLDN7* in breast (Fig. 6D, left; fig. S9C) was identified as an epithelial cell ieQTL in breast (*19*), and both the sb-eQTL and cell type ieQTL signal colocalize with the birth weight GWAS signal (fig. S9D). These results suggest that gene-by-environment (G×E) interactions – with a broad definition of the environment (sex, in the present study) – detected at eQTLs may derive from eQTLs that are specific to cell types correlated with the studied environment. While these are not false positive signals *per se*, taking cell type composition into account in the interpretation of eQTLs interacting with biological variables, is essential. Importantly, we have demonstrated that characterization of eQTL-related G×Sex interactions can provide mechanistic insights on the sex-differentiated genetic architecture of complex traits.

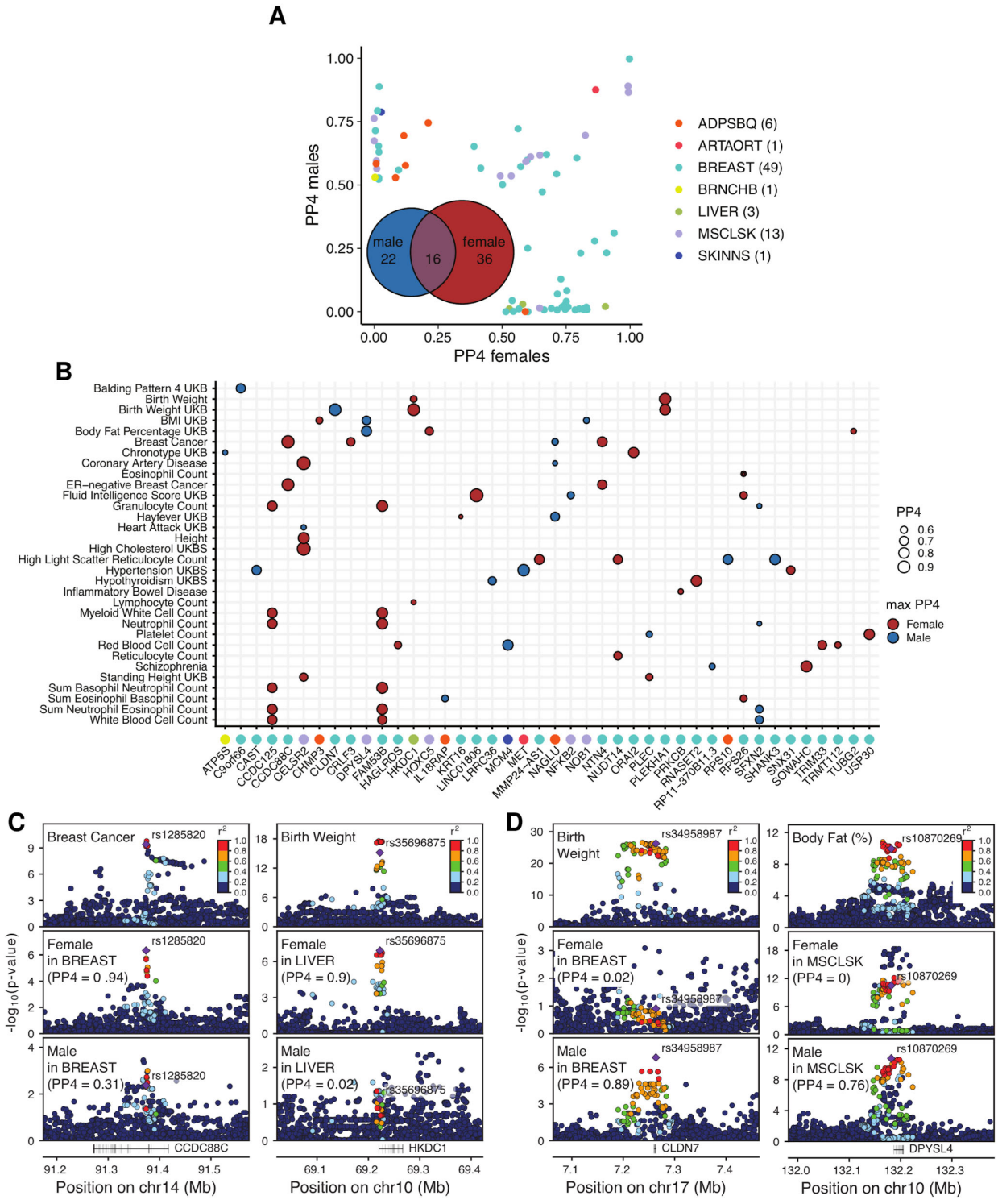


Fig. 6. Colocalization of sex-biased eQTLs (sb-eQTLs) with GWAS signal (**A**) Posterior probability (PP4) of 74 colocalized gene-trait pairs where a GWAS shows evidence of colocalization with the female- and/or the male-stratified *cis*-eQTL signal (PP4 > 0.5). Inset: number of colocalizing loci for female and male eQTLs. Legend: in parenthesis, number of colocalizing loci per tissue. (**B**) GWAS-eQTL colocalizing genes (PP4 > 0.5) color-labeled

by eQTL tissue of origin according to labels in Fig. 1A (x-axis), are categorized by the sex where the colocalization signal is maximized with the corresponding GWAS trait (y-axis). Comparing the colocalization PP4 values for male and female *cis*-eQTL signal, the estimates can be maximum in females (red) or males (blue). (**C**) Genotype-phenotype p-value landscapes of the *CCDC88C* (left panel) and *HKDC1* (right panel) locus. For the *CCDC88C* locus, panels illustrate GWAS signal for breast cancer (top) and *CCDC88C cis*-eQTL signal for females (middle) and males (bottom) in breast mammary tissue. For the *HKDC1* locus, panels illustrate GWAS signal for birth weight (top) and *HKDC1 cis*-eQTL signal for females (middle) and males (bottom) in liver. (**D**) Genotype-phenotype p-value landscapes of the *CLDN7* (left panel) and *DPYSL4* (right panel) locus. For the *CLDN7* locus, panels illustrate GWAS signal for birth weight (top) and *CLDN7 cis*-eQTL signal for females (middle) and males (bottom) in breast mammary tissue. For the *DPYSL4* locus, panels illustrate GWAS signal for body fat (top) and *DPYSL4 cis*-eQTL signal for females (middle) and males (bottom) in muscle skeletal tissue.

Discussion

We have performed a comprehensive characterization of sex-biased gene expression and sex differences in the genetic regulation of gene expression in *cis* across 44 human tissue sources. We also characterized sex differences in tissue cell type composition for seven cell types. Together, these results comprise the largest resources to date characterizing sex differences in the human transcriptome.

We report widespread sex-biased gene expression in all tissues (as in (*9*, *52–55*)), with 37% of genes exhibiting sex bias in at least one tissue, representing diverse molecular and biological functions, and including genes relevant to disease and clinical phenotypes. The strongest sex bias was observed for X-chromosome genes, as expected due to escape from X-inactivation in females. Our analysis identified and characterized cross-tissue expression of novel candidate 'escape genes' while confirming others that were previously characterized in 29 GTEx tissues (*13*). The vast majority of sex-biased genes were autosomal, suggesting the influence of sex on genome-wide regulatory programs. Moreover, we discovered that a portion of these genes were non-randomly distributed across the genome, suggesting sex differences in regional regulation, which has been reported (*16*) but not well characterized to date. The impact of sex on expression of individual genes was generally small, deriving from highly overlapping male and female distributions of interindividual variation, indicative of differential expression as opposed to fully dimorphic expression.

Although we identified a set of X-linked genes with sex-biased expression across many tissues, the overall tissue-sharing of sex-biased expression was strongly skewed toward tissue-specificity,

with 23.8% of sex-biased genes discovered in only a single tissue. This specificity was not driven by overall gene expression patterns across tissues, suggesting highly tissue-specific control of sex-biased gene regulation. Analysis of transcription factor binding sites in promoters of sex-biased genes revealed tissue-specific enrichments, implicating specific transcription factors in mediating sex-differentiated transcriptional profiles within and across tissues. The analysis also yielded the expected enrichment for hormone-related transcription factors that were shared more broadly across tissues, as well as downstream effects from X-inactivation escapees.

Sex differences in cell type abundances can induce sex-biased expression, where differences in expression profiles of different cell types contribute to composite measures of expression in tissues, impacting interpretation. The differential expression analysis approach employed here controlled for known and unknown confounders that partially account for inter-individual variation in cell type composition within tissues, thus our results are less impacted by sex differences in tissue cellular composition than previous studies (*9*, *55*, *56*). More recent studies (e.g., (*52*)) have implemented analysis approaches to at least partially account for this issue.

In contrast to the strong impact of sex on gene expression levels, the overall extent of sex effects on genetic regulation in *cis* is much less (369 sb-eQTLs at FDR ≤ 0.25). Our analysis approach tested for genotype-by-sex interaction for *cis*-eQTLs that had been identified in the combined-sex *cis*-eQTL GTEx analysis (*31*), ultimately discovering that a small portion (0.07%) of those *cis*-eQTLs are driven in part (or primarily) by genetic effects in one sex.

The difference in the extent of sex-biased expression and sb-eQTL discovery can be attributed, in part, through differences in power of the two analyses. For sb-eQTLs, the combination of small genotype-by-sex interaction effect sizes, high inter-individual expression heterogeneity, and the sex imbalance in the GTEx collection affect the power of the interaction test. This suggests that to better characterize this phenomenon in the future, much larger cohorts are needed, particularly to assess sex effects for all *cis*-variants and all genes. The relatively modest number of interactions for a factor as impactful as sex also suggests that for other, more subtle environmental factors, interactions are likely to be challenging to identify (as noted in (*34*)). Similarly to the differential gene expression analysis, the sb-eQTL analysis is impacted by cell type heterogeneity within tissues. Indeed, we demonstrate that a portion of sb-eQTLs are mediated by cell type composition,

suggesting that a portion of the sb-eQTL signal may derive from the combination of cell type-specific eQTLs and sex differences in the tissue's cell type composition. The remaining loci for which we had no evidence of cell type mediation may represent true sex differences in genetic regulation of these genes, but might also derive from unknown factors confounded with sex, including cell types that were not part of our analysis; thus the full impact of cell type differences across tissues remains to be determined. Future sb-eQTL analysis of single cell types will help disentangle sex effects on the genetics of gene expression that derive from sex differences in tissue composition versus those that derive from sex chromosome status. However, this approach also has limitations due to the removal of cells from the *in situ* tissue environment, including, for example, the presence of other cell types and hormonal environment. Regardless of the sex bias source, we have discovered and characterized the largest collection of human tissue sb-eQTLs to date.

To understand the molecular basis of sex differences in disease, it will be important to investigate sex differences at the molecular level in patients and to assess those differences in the context of non-diseased individuals. For example, some cancers exhibit strong sex differences in prevalence and other features (*1*) and also exhibit sex-biased gene expression in tumors (*57*). The results presented here can serve as a resource of sex effects in 'normal' tissues to compare with disease cohorts. Beyond gene expression, sex-biased genetic regulation may also contribute to higher order phenotypes such as complex traits and diseases; colocalization analysis of sex-stratified *cis*-eQTLs and GWAS summary statistics yielded novel variant-gene-trait associations that were not found using combined-sex *cis*-eQTLs either due to allelic heterogeneity in the combined-sex cohort or due to genetic effects on gene expression that are (predominantly) driven by a single sex. We have therefore demonstrated that sex-aware colocalization analyses can provide insights into the sex-differentiated genetic architecture of disease. Future work in this area will be to combine sex-stratified *cis*-eQTLs with summary statistics from sex-stratified GWAS to fully comprehend the impact of sex on human health and disease.

References

1. D. Zheng, J. Trynda, C. Williams, J. A. Vold, J. H. Nguyen, D. M. Harnois, S. P. Bagaria, S. A. McLaughlin, Z. Li, Sexual dimorphism in the incidence of human cancers.

BMC Cancer. **19**, 684 (2019).

2. G. D. Anderson, Sex and racial differences in pharmacological response: where is the evidence? Pharmacogenetics, pharmacokinetics, and pharmacodynamics. *J. Womens. Health* . **14**, 19–29 (2005).

3. V. Kuan, S. Denaxas, A. Gonzalez-Izquierdo, K. Direk, O. Bhatti, S. Husain, S. Sutaria, M. Hingorani, D. Nitsch, C. A. Parisinos, R. T. Lumbers, R. Mathur, R. Sofat, J. P. Casas, I. C. K. Wong, H. Hemingway, A. D. Hingorani, A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. *Lancet Digit Health*. **1**, e63–e77 (2019).

4. S. T. Ngo, F. J. Steyn, P. A. McCombe, Gender differences in autoimmune disease. *Front. Neuroendocrinol.* **35**, 347–369 (2014).

5. D. Westergaard, P. Moseley, F. K. H. Sørup, P. Baldi, S. Brunak, Population-wide analysis of differences in disease progression patterns in men and women. *Nat. Commun.* **10**, 1–14 (2019).

6. E. A. Khramtsova, L. K. Davis, B. E. Stranger, The role of sex in the genomics of human complex traits. *Nat. Rev. Genet.* **20**, 173–190 (2019).

7. C. W. Law, Y. Chen, W. Shi, G. K. Smyth, voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).

8. S. M. Urbut, G. Wang, P. Carbonetto, M. Stephens, Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019).

9. M. Melé, P. G. Ferreira, F. Reverter, D. S. DeLuca, J. Monlong, M. Sammeth, T. R. Young, J. M. Goldmann, D. D. Pervouchine, T. J. Sullivan, R. Johnson, A. V. Segrè, S. Djebali, A. Niarchou, GTEx Consortium, F. A. Wright, T. Lappalainen, M. Calvo, G. Getz, E. T. Dermitzakis, K. G. Ardlie, R. Guigó, Human genomics. The human transcriptome across tissues and individuals. *Science*. **348**, 660–665 (2015).

10. S. E. Ellis, L. Collado-Torres, A. Jaffe, J. T. Leek, Improving the value of public RNA-seq expression data by phenotype prediction. *Nucleic Acids Res.* **46**, e54–e54 (2018).

11. B. P. Balaton, C. J. Brown, Escape Artists of the X Chromosome. *Trends Genet.* **32**, 348–359 (2016).

12. K. De Preter, R. Barriot, F. Speleman, J. Vandesompele, Y. Moreau, Positional gene enrichment analysis of gene sets for high-resolution identification of overrepresented chromosomal regions. *Nucleic Acids Res.* **36**, e43 (2008).

13. T. Tukiainen, A.-C. Villani, A. Yen, M. A. Rivas, J. L. Marshall, R. Satija, M. Aguirre, L. Gauthier, M. Fleharty, A. Kirby, B. B. Cummings, S. E. Castel, K. J. Karczewski, F. Aguet, A. Byrnes, GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration &Visualization—EBI, Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz, T. Lappalainen, A. Regev, K. G. Ardlie, N. Hacohen, D. G. MacArthur, Landscape of X chromosome inactivation across human tissues. *Nature*. **550**, 244–248 (2017).

14. M. Gheorghe, G. K. Sandve, A. Khan, J. Chèneby, B. Ballester, A. Mathelier, A map of direct TF–DNA interactions in the human genome. *Nucleic Acids Res.* **47**, e21–e21 (2019).

15. E. Fiorito, Y. Sharma, S. Gilfillan, S. Wang, S. K. Singh, S. V. Satheesh, M. R. Katika, A. Urbanucci, B. Thiede, I. G. Mills, A. Hurtado, CTCF modulates Estrogen Receptor function through specific chromatin and nuclear matrix interactions. *Nucleic Acids Res.* **44**, 10588–10602 (2016).

16. B. J. Matthews, D. J. Waxman, CTCF and Cohesin link sex-biased distal regulatory elements to sex-biased gene expression in mouse liver. *bioRxiv* (2019) (available at https://www.biorxiv.org/content/10.1101/577577v1.abstract).

17. N. Brockdorff, Polycomb complexes in X chromosome inactivation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372** (2017), doi:10.1098/rstb.2017.0021.

18. D. Lau-Corona, W. K. Bae, L. Hennighausen, D. J. Waxman, Sex-biased genetic programs in liver metabolism and liver fibrosis are controlled by EZH1 and EZH2. *bioRxiv*, 577056 (2019).

19. S. Kim-Hellmuth, F. Aguet, M. Oliva, M. Muñoz-Aguirre, V. Wucher, S. Kasela, S. E. Castel, A. R. Hamel, A. Viñuela, A. L. Roberts, S. Mangul, X. Wen, G. Wang, A. N. Barbeira, D. Garrido-Martín, B. Nadel, Y. Zou, R. Bonazzola, J. Quan, A. Brown, A. Martinez-Perez, J. M. Soria, G. Consortium, G. Getz, E. T. Dermitzakis, K. S. Small, M. Stephens, H. S. Xi, H. K. Im, R. Guigó, A. V. Segrè, B. E. Stranger, K. G. Ardlie, T. Lappalainen, Cell type specific genetic regulation of gene expression across human tissues. *bioRxiv*, 806117 (2019).

20. Y. Chen, Y. Zhang, G. Zhao, C. Chen, P. Yang, S. Ye, X. Tan, Difference in Leukocyte Composition between Women before and after Menopausal Age, and Distinct Sexual Dimorphism. *PLoS One*. **11**, e0162953–e0162953 (2016).

21. E. Bongen, H. Lucian, A. Khatri, G. K. Fragiadakis, Z. B. Bjornson, G. P. Nolan, P. J. Utz, P. Khatri, Sex Differences in the Blood Transcriptome Identify Robust Changes in Immune Cell Proportions with Aging and Influenza Infection. *Cell Rep.* **29**, 1961–1973.e4 (2019).

22. Alessandra Breschi, Manuel Muñoz-Aguirre, Valentin Wucher, Carrie A. Davis, Diego Garrido-Martín, Sarah Djebali, Jesse Gillis, Dmitri D. Pervouchine, Anna Vlasova, Alexander Dobin, Chris Zaleski, Jorg Drenkow, Cassidy Danyko, Alexandra Scavelli, Ferran Reverter, Michael P. Snyder, Thomas R. Gingeras, Roderic Guigó, A limited set of transcriptional programs define major histological types and provide the molecular basis for

a cellular taxonomy of the human body. *ASHG Abstract* (2018) (available at https://eventpilot.us/web/page.php?page=IntHtml&project=ASHG18&id=180120926).

23. H. E. Carlson, Approach to the Patient with Gynecomastia. *None*. **96**, 15–21 (2011).

24. K. Abe, N. Matsuura, Y. Nohara, H. Fujita, K. Fujieda, T. Kato, Y. Mikami, Prolactin response to thyrotropin-releasing hormone in children with gynecomastia, premature thelarche and idiopathic precocious puberty. *Tohoku J. Exp. Med.* **142**, 283–288 (1984).

25. C. Giordano, G. Stassi, R. D. Maria, M. Todaro, P. Richiusa, G. Papoff, G. Ruberti, M. Bagnasco, R. Testi, A. Galluzzo, Potential Involvement of Fas and Its Ligand in the Pathogenesis of Hashimoto's Thyroiditis. *Science*. **275**, 960–963 (1997).

26. R. Parker, The role of adipose tissue in fatty liver diseases. *Liver Research*. **2**, 35–42 (2018).

27. K. Rawlik, O. Canela-Xandri, A. Tenesa, Evidence for sex-specific genetic architectures across a spectrum of human complex traits. *Genome Biol.* **17**, 166 (2016).

28. E. Flynn, Y. Tanigawa, F. Rodriguez, R. B. Altman, N. Sinnott-Armstrong, M. A. Rivas, Sex-specific genetic effects across biomarkers. *bioRxiv* (2019), p. 837021.

29. D. Shungin, T. W. Winkler, D. C. Croteau-Chonka, T. Ferreira, A. E. Locke, R. Mägi, R. J. Strawbridge, T. H. Pers, K. Fischer, A. E. Justice, T. Workalemahu, J. M. W. Wu, M. L. Buchkovich, N. L. Heard-Costa, T. S. Roman, A. W. Drong, C. Song, S. Gustafsson, F. R. Day, T. Esko, T. Fall, Z. Kutalik, J. Luan, J. C. Randall, A. Scherag, S. Vedantam, A. R. Wood, J. Chen, R. Fehrmann, J. Karjalainen, B. Kahali, C.-T. Liu, E. M. Schmidt, D. Absher, N. Amin, D. Anderson, M. Beekman, J. L. Bragg-Gresham, S. Buyske, A. Demirkan, G. B. Ehret, M. F. Feitosa, A. Goel, A. U. Jackson, T. Johnson, M. E. Kleber, K. Kristiansson, M. Mangino, I. Mateo Leach, C. Medina-Gomez, C. D. Palmer, D. Pasko, S. Pechlivanis, M. J. Peters, I. Prokopenko, A. Stančáková, Y. Ju Sung, T. Tanaka, A. Teumer, J. V. Van Vliet-Ostaptchouk, L. Yengo, W. Zhang, E. Albrecht, J. Ärnlöv, G. M.

Arscott, S. Bandinelli, A. Barrett, C. Bellis, A. J. Bennett, C. Berne, M. Blüher, S. Böhringer, F. Bonnet, Y. Böttcher, M. Bruinenberg, D. B. Carba, I. H. Caspersen, R. Clarke, E. Warwick Daw, J. Deelen, E. Deelman, G. Delgado, A. S. F. Doney, N. Eklund, M. R. Erdos, K. Estrada, E. Eury, N. Friedrich, M. E. Garcia, V. Giedraitis, B. Gigante, A. S. Go, A. Golay, H. Grallert, T. B. Grammer, J. Gräßler, J. Grewal, C. J. Groves, T. Haller, G. Hallmans, C. A. Hartman, M. Hassinen, C. Hayward, K. Heikkilä, K.-H. Herzig, Q. Helmer, H. L. Hillege, O. Holmen, S. C. Hunt, A. Isaacs, T. Ittermann, A. L. James, I. Johansson, T. Juliusdottir, I.-P. Kalafati, L. Kinnunen, W. Koenig, I. K. Kooner, W. Kratzer, C. Lamina, K. Leander, N. R. Lee, P. Lichtner, L. Lind, J. Lindström, S. Lobbens, M. Lorentzon, F. Mach, P. K. E. Magnusson, A. Mahajan, W. L. McArdle, C. Menni, S. Merger, E. Mihailov, L. Milani, R. Mills, A. Moayyeri, K. L. Monda, S. P. Mooijaart, T. W. Mühleisen, A. Mulas, G. Müller, M. Müller-Nurasyid, R. Nagaraja, M. A. Nalls, N. Narisu, N. Glorioso, I. M. Nolte, M. Olden, N. W. Rayner, F. Renstrom, J. S. Ried, N. R. Robertson, L. M. Rose, S. Sanna, H. Scharnagl, S. Scholtens, B. Sennblad, T. Seufferlein, C. M. Sitlani, A. Vernon Smith, K. Stirrups, H. M. Stringham, J. Sundström, M. A. Swertz, A. J. Swift, A.-C. Syvänen, B. O. Tayo, B. Thorand, G. Thorleifsson, A. Tomaschitz, C. Troffa, F. V. A. van Oort, N. Verweij, J. M. Vonk, L. L. Waite, R. Wennauer, T. Wilsgaard, M. K. Wojczynski, A. Wong, Q. Zhang, J. Hua Zhao, E. P. Brennan, M. Choi, P. Eriksson, L. Folkersen, A. Franco-Cereceda, A. G. Gharavi, Å. K. Hedman, M.-F. Hivert, J. Huang, S. Kanoni, F. Karpe, S. Keildson, K. Kiryluk, L. Liang, R. P. Lifton, B. Ma, A. J. McKnight, R. McPherson, A. Metspalu, J. L. Min, M. F. Moffatt, G. W. Montgomery, J. M. Murabito, G. Nicholson, D. R. Nyholt, C. Olsson, J. R. B. Perry, E. Reinmaa, R. M. Salem, N. Sandholm, E. E. Schadt, R. A. Scott, L. Stolk, E. E. Vallejo, H.-J. Westra, K. T. Zondervan, The ADIPOGen Consortium, The CARDIOGRAMplusC4D Consortium, The CKDGen Consortium, The GEFOS Consortium, The GENIE Consortium, The Glgc, The Icbp, The International Endogene Consortium, The LifeLines Cohort Study, The MAGIC Investigators, The MuTHER Consortium, The PAGE Consortium, The ReproGen Consortium, P. Amouyel, D. Arveiler, S. J. L. Bakker, J. Beilby, R. N. Bergman, J. Blangero, M. J. Brown, M. Burnier, H. Campbell, A. Chakravarti, P. S. Chines, S. Claudi-Boehm, F. S. Collins, D. C. Crawford, J. Danesh, U. de Faire, E. J. C. de Geus, M. Dörr, R. Erbel, J. G. Eriksson, M. Farrall, E. Ferrannini, J. Ferrières, N. G. Forouhi, T. Forrester, O.

H. Franco, R. T. Gansevoort, C. Gieger, V. Gudnason, C. A. Haiman, T. B. Harris, A. T. Hattersley, M. Heliövaara, A. A. Hicks, A. D. Hingorani, W. Hoffmann, A. Hofman, G. Homuth, S. E. Humphries, E. Hyppönen, T. Illig, M.-R. Jarvelin, B. Johansen, P. Jousilahti, A. M. Jula, J. Kaprio, F. Kee, S. M. Keinanen-Kiukaanniemi, J. S. Kooner, C. Kooperberg, P. Kovacs, A. T. Kraja, M. Kumari, K. Kuulasmaa, J. Kuusisto, T. A. Lakka, C. Langenberg, L. Le Marchand, T. Lehtimäki, V. Lyssenko, S. Männistö, A. Marette, T. C. Matise, C. A. McKenzie, B. McKnight, A. W. Musk, S. Möhlenkamp, A. D. Morris, M. Nelis, C. Ohlsson, A. J. Oldehinkel, K. K. Ong, L. J. Palmer, B. W. Penninx, A. Peters, P. P. Pramstaller, O. T. Raitakari, T. Rankinen, D. C. Rao, T. K. Rice, P. M. Ridker, M. D. Ritchie, I. Rudan, V. Salomaa, N. J. Samani, J. Saramies, M. A. Sarzynski, P. E. H. Schwarz, A. R. Shuldiner, J. A. Staessen, V. Steinthorsdottir, R. P. Stolk, K. Strauch, A. Tönjes, A. Tremblay, E. Tremoli, M.-C. Vohl, U. Völker, P. Vollenweider, J. F. Wilson, J. C. Witteman, L. S. Adair, M. Bochud, B. O. Boehm, S. R. Bornstein, C. Bouchard, S. Cauchi, M. J. Caulfield, J. C. Chambers, D. I. Chasman, R. S. Cooper, G. Dedoussis, L. Ferrucci, P. Froguel, H.-J. Grabe, A. Hamsten, J. Hui, K. Hveem, K.-H. Jöckel, M. Kivimaki, D. Kuh, M. Laakso, Y. Liu, W. März, P. B. Munroe, I. Njølstad, B. A. Oostra, C. N. A. Palmer, N. L. Pedersen, M. Perola, L. Pérusse, U. Peters, C. Power, T. Quertermous, R. Rauramaa, F. Rivadeneira, T. E. Saaristo, D. Saleheen, J. Sinisalo, P. Eline Slagboom, H. Snieder, T. D. Spector, U. Thorsteinsdottir, M. Stumvoll, J. Tuomilehto, A. G. Uitterlinden, M. Uusitupa, P. van der Harst, G. Veronesi, M. Walker, N. J. Wareham, H. Watkins, H.-E. Wichmann, G. R. Abecasis, T. L. Assimes, S. I. Berndt, M. Boehnke, I. B. Borecki, P. Deloukas, L. Franke, T. M. Frayling, L. C. Groop, D. J. Hunter, R. C. Kaplan, J. R. O'Connell, L. Qi, D. Schlessinger, D. P. Strachan, K. Stefansson, C. M. van Duijn, C. J. Willer, P. M. Visscher, J. Yang, J. N. Hirschhorn, M. Carola Zillikens, M. I. McCarthy, E. K. Speliotes, K. E. North, C. S. Fox, I. Barroso, P. W. Franks, E. Ingelsson, I. M. Heid, R. J. F. Loos, L. A. Cupples, A. P. Morris, C. M. Lindgren, K. L. Mohlke, New genetic loci link adipose and insulin biology to body fat distribution. *Nature*. **518**, 187–196 (2015).

30. S. L. Pulit, C. Stoneman, A. P. Morris, A. R. Wood, C. A. Glastonbury, J. Tyrrell, L. Yengo, T. Ferreira, E. Marouli, Y. Ji, J. Yang, S. Jones, R. Beaumont, D. C. Croteau-Chonka, T. W. Winkler, G. Consortium, A. T. Hattersley, R. J. F. Loos, J. N. Hirschhorn, P.

M. Visscher, T. M. Frayling, H. Yaghootkar, C. M. Lindgren, Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).

31. F. Aguet, A. N. Barbeira, R. Bonazzola, A. Brown, S. E. Castel, B. Jo, S. Kasela, S. Kim-Hellmuth, Y. Liang, M. Oliva, P. E. Parsana, E. Flynn, L. Fresard, E. R. Gaamzon, A. R. Hamel, Y. He, F. Hormozdiari, P. Mohammadi, M. Muñoz-Aguirre, Y. Park, A. Saha, A. V. Segrć, B. J. Strober, X. Wen, V. Wucher, S. Das, D. Garrido-Martín, N. R. Gay, R. E. Handsaker, P. J. Hoffman, S. Kashin, A. Kwong, X. Li, D. MacArthur, J. M. Rouhana, M. Stephens, E. Todres, A. Viñuela, G. Wang, Y. Zou, T. G. Consortium, C. D. Brown, N. Cox, E. Dermitzakis, B. E. Engelhardt, G. Getz, R. Guigo, S. B. Montgomery, B. E. Stranger, H. K. Im, A. Battle, K. G. Ardlie, T. Lappalainen, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *bioRxiv*, 787903 (2019).

32. H. Ongen, A. Buil, A. A. Brown, E. T. Dermitzakis, O. Delaneau, Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics*. **32**, 1479–1485 (2016).

33. S. E. Castel, F. Aguet, P. Mohammadi, G. Consortium, K. G. Ardlie, T. Lappalainen, A vast resource of allelic expression data spanning human tissues. *bioRxiv*, 792911 (2019).

34. D. A. Knowles, J. R. Davis, H. Edgington, A. Raj, M.-J. Favé, X. Zhu, J. B. Potash, M. M. Weissman, J. Shi, D. F. Levinson, P. Awadalla, S. Mostafavi, S. B. Montgomery, A. Battle, Allele-specific expression reveals interactions between genetic variation and environment. *Nat. Methods*. **14**, 699–702 (2017).

35. J. J. Shen, Y.-F. Wang, W. Yang, Sex-Interacting mRNA- and miRNA-eQTLs and Their Implications in Gene Expression Regulation and Disease. *Front. Genet.* **10** (2019), doi:10.3389/fgene.2019.00313.

36. K. R. Kukurba, P. Parsana, B. Balliu, K. S. Smith, Z. Zappala, D. A. Knowles, M.-J. Favé, J. R. Davis, X. Li, X. Zhu, J. B. Potash, M. M. Weissman, J. Shi, A. Kundaje,

D. F. Levinson, P. Awadalla, S. Mostafavi, A. Battle, S. B. Montgomery, Impact of the X Chromosome and sex on regulatory variation. *Genome Res.*, gr.197897.115 (2016).

37. C. Yao, R. Joehanes, A. D. Johnson, T. Huan, T. Esko, S. Ying, J. E. Freedman, J. Murabito, K. L. Lunetta, A. Metspalu, P. J. Munson, D. Levy, Sex- and age-interacting eQTLs in human complex diseases. *Hum. Mol. Genet.* **23**, 1947–1956 (2014).

38. The GAIT2 project, (available at http://ugcd.github.io/pages/projects/gait2/).

39. A. C. Leon, M. Heo, Sample Sizes Required to Detect Interactions between Two Binary Fixed-Effects in a Mixed-Effects Linear Regression Model. *Comput. Stat. Data Anal.* **53**, 603–608 (2009).

40. A. Tang, K. Gao, L. Chu, R. Zhang, J. Yang, J. Zheng, Aurora kinases: novel therapy targets in cancers. *Oncotarget*. **8**, 23937–23954 (2017).

41. K. Dhanasekaran, A. Bose, V. J. Rao, R. Boopathi, S. R. Shankar, V. K. Rao, A. Swaminathan, M. Vasudevan, R. Taneja, T. K. Kundu, Unraveling the role of aurora A beyond centrosomes and spindle assembly: implications in muscle differentiation. *FASEB J.* **33**, 219–230 (2019).

42. C. Giambartolomei, D. Vukcevic, E. E. Schadt, L. Franke, A. D. Hingorani, C. Wallace, V. Plagnol, Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).

43. T. Zhan, N. Rindtorff, M. Boutros, Wnt signaling in cancer. *Oncogene*. **36**, 1461–1473 (2017).

44. X. Xu, Q. Yan, Y. Wang, X. Dong, NTN4 is associated with breast cancer metastasis via regulation of EMT-related biomarkers. *Oncol. Rep.* **37**, 449–457 (2017).

45. M. M. Marjaneh, H. Sivakumaran, K. M. Hillman, S. Kaufmann, N. Hussein, L. G. Lima, S. Ham, S. Kar, J. Beesley, L. Fachal, D. F. Easton, A. M. Dunning, A. Möller, G. Chenevix-Trench, S. L. Edwards, J. D. French, High-throughput allelic expression

imbalance analyses identify candidate breast cancer risk genes. *bioRxiv*, 521013 (2019).

46. J. D. Betteridge, S. P. Armbruster, C. Maydonovitch, G. R. Veerappan, Inflammatory bowel disease prevalence by age, gender, race, and geographic location in the U.S. military health care population. *Inflamm. Bowel Dis.* **19**, 1421–1427 (2013).

47. T. Strunz, F. Grassmann, J. Gayán, S. Nahkuri, D. Souza-Costa, C. Maugeais, S. Fauser, E. Nogoceke, B. H. F. Weber, A mega-analysis of expression quantitative trait loci (eQTL) provides insight into the regulatory architecture of gene expression variation in liver. *Sci. Rep.* **8** (2018), doi:10.1038/s41598-018-24219-z.

48. M. G. Hayes, M. Urbanek, M.-F. Hivert, L. L. Armstrong, J. Morrison, C. Guo, L. P. Lowe, D. A. Scheftner, A. Pluzhnikov, D. M. Levine, C. P. McHugh, C. M. Ackerman, L. Bouchard, D. Brisson, B. T. Layden, D. Mirel, K. F. Doheny, M. V. Leya, R. N. Lown-Hecht, A. R. Dyer, B. E. Metzger, T. E. Reddy, N. J. Cox, W. L. Lowe, HAPO Study Cooperative Research Group, Identification of HKDC1 and BACE2 as genes influencing glycemic traits during pregnancy through genome-wide association studies. *Diabetes*. **62**, 3282–3291 (2013).

49. W. L. Lowe, D. M. Scholtens, V. Sandler, M. G. Hayes, Genetics of Gestational Diabetes Mellitus and Maternal Metabolism. *Curr. Diab. Rep.* **16**, 15 (2016).

50. C. Guo, A. E. Ludvik, M. E. Arlotto, M. G. Hayes, L. L. Armstrong, D. M. Scholtens, C. D. Brown, C. B. Newgard, T. C. Becker, B. T. Layden, W. L. Lowe, T. E. Reddy, Coordinated regulatory variation associated with gestational hyperglycaemia regulates expression of the novel hexokinase HKDC1. *Nat. Commun.* **6**, 6069 (2015).

51. H. Nagano, N. Hashimoto, A. Nakayama, S. Suzuki, Y. Miyabayashi, A. Yamato, S. Higuchi, M. Fujimoto, I. Sakuma, M. Beppu, M. Yokoyama, Y. Suzuki, S. Sugano, K. Ikeda, I. Tatsuno, I. Manabe, K. Yokote, S. Inoue, T. Tanaka, p53-inducible DPYSL4 associates with mitochondrial supercomplexes and regulates energy metabolism in adipocytes and cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 8370–8375 (2018).

52. S. Naqvi, A. K. Godfrey, J. F. Hughes, M. L. Goodheart, R. N. Mitchell, D. C.

Page, Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science*. **365** (2019), doi:10.1126/science.aaw7317.

53. I. Kassam, Y. Wu, J. Yang, P. M. Visscher, A. F. McRae, Tissue-specific sex differences in human gene expression. *Hum. Mol. Genet.* **28**, 2976–2986 (2019).

54. M. Gershoni, S. Pietrokovski, The landscape of sex-differential transcriptome and its consequent selection in human adults. *BMC Biol.* **15**, 7 (2017).

55. B. T. Mayne, T. Bianco-Miotto, S. Buckberry, J. Breen, V. Clifton, C. Shoubridge, C. T. Roberts, Large Scale Gene Expression Meta-Analysis Reveals Tissue-Specific, Sex-Biased Gene Expression in Humans. *Front. Genet.* **7**, 183 (2016).

56. D. Trabzuni, A. Ramasamy, S. Imran, R. Walker, C. Smith, M. E. Weale, J. Hardy, M. Ryten, Widespread sex differences in gene expression and splicing in the adult human brain. *Nat. Commun.* **4**, 1–7 (2013).

57. Y. Yuan, L. Liu, H. Chen, Y. Wang, Y. Xu, H. Mao, J. Li, G. B. Mills, Y. Shu, L. Li, H. Liang, Comprehensive Characterization of Molecular Differences in Cancer between Male and Female Patients. *Cancer Cell*. **29**, 711–722 (2016).

58. D. Aran, Z. Hu, A. J. Butte, xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).

59. J. Chen, E. Behnam, J. Huang, M. F. Moffatt, D. J. Schaid, L. Liang, X. Lin, Fast and robust adjustment of cell mixtures in epigenome-wide association studies with SmartSVA. *BMC Genomics*. **18**, 413 (2017).

60. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9440–9445 (2003).

61. M. Stephens, False discovery rates: a new deal. *Biostatistics*. **18**, 275–294 (2017).

62. N. Kryuchkova-Mostacci, M. Robinson-Rechavi, A benchmark of gene expression tissue-specificity metrics. *Brief. Bioinform.* **18**, 205–214 (2017).

63. R. Suzuki, H. Shimodaira, *pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling* (2015).

64. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2012).

65. T. Chen, C. Guestrin, in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, NY, USA, 2016), *KDD '16*, pp. 785–794.

66. S. M. Lundberg, S.-I. Lee, in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett, Eds. (Curran Associates, Inc., 2017), pp. 4765–4774.

67. B. J. Schmiedel, D. Singh, A. Madrigal, A. G. Valdovino-Gonzalez, B. M. White, J. Zapardiel-Gonzalo, B. Ha, G. Altay, J. A. Greenbaum, G. McVicker, G. Seumois, A. Rao, M. Kronenberg, B. Peters, P. Vijayanand, Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell*. **175**, 1701–1715.e16 (2018).

68. N. C. Sheffield, C. Bock, LOLA: Enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* (2016) (available at http://code.databio.org/LOLA).

69. A. A. Sergushichev, An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv*, 060012 (2016).

70. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).

71. G. Su, A. Kuchinsky, J. H. Morris, D. J. States, F. Meng, GLay: community structure analysis of biological networks. *Bioinformatics*. **26**, 3135–3137 (2010).

72. W. J. Astle, H. Elding, T. Jiang, D. Allen, D. Ruklisa, A. L. Mann, D. Mead, H. Bouman, F. Riveros-Mckay, M. A. Kostadima, J. J. Lambourne, S. Sivapalaratnam, K. Downes, K. Kundu, L. Bomba, K. Berentsen, J. R. Bradley, L. C. Daugherty, O. Delaneau, K. Freson, S. F. Garner, L. Grassi, J. Guerrero, M. Haimel, E. M. Janssen-Megens, A. Kaan, M. Kamat, B. Kim, A. Mandoli, J. Marchini, J. H. A. Martens, S. Meacham, K. Megy, J. O'Connell, R. Petersen, N. Sharifi, S. M. Sheard, J. R. Staley, S. Tuna, M. van der Ent, K. Walter, S.-Y. Wang, E. Wheeler, S. P. Wilder, V. Iotchkova, C. Moore, J. Sambrook, H. G. Stunnenberg, E. Di Angelantonio, S. Kaptoge, T. W. Kuijpers, E. Carrillo-de-Santa-Pau, D. Juan, D. Rico, A. Valencia, L. Chen, B. Ge, L. Vasquez, T. Kwan, D. Garrido-Martín, S. Watt, Y. Yang, R. Guigo, S. Beck, D. S. Paul, T. Pastinen, D. Bujold, G. Bourque, M. Frontini, J. Danesh, D. J. Roberts, W. H. Ouwehand, A. S. Butterworth, N. Soranzo, The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell*. **167**, 1415–1429.e19 (2016).

73. O. Stegle, L. Parts, M. Piipari, J. Winn, R. Durbin, Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7**, 500–507 (2012).

74. Y. Nédélec, J. Sanz, G. Baharian, Z. A. Szpiech, A. Pacis, A. Dumaine, J.-C. Grenier, A. Freiman, A. J. Sams, S. Hebert, A. Pagé Sabourin, F. Luca, R. Blekhman, R. D. Hernandez, R. Pique-Regi, J. Tung, V. Yotova, L. B. Barreiro, Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell*. **167**, 657–669.e21 (2016).

75. P. Mohammadi, S. E. Castel, A. A. Brown, T. Lappalainen, Quantifying the regulatory effect size of cis-acting genetic variation using allelic fold change. *Genome Res.* **27**, 1872–1884 (2017).

76. S. E. Castel, P. Mohammadi, W. K. Chung, Y. Shen, T. Lappalainen, Rare variant phasing and haplotypic expression from RNA sequencing with phASER. *Nat. Commun.* **7**, 12817 (2016).

77. D. A. Knowles, C. K. Burrows, J. D. Blischak, K. M. Patterson, D. J. Serie, N.

Norton, C. Ober, J. K. Pritchard, Y. Gilad, Determining the genetic basis of anthracycline-cardiotoxicity by molecular response QTL mapping in induced cardiomyocytes. *Elife*. **7** (2018), doi:10.7554/eLife.33480.

78. A. Saha, A. Battle, False positives in trans-eQTL and co-expression analyses arising from RNA-sequencing alignment errors. *F1000Res.* **7**, 1860 (2018).

79. M. Pividori, P. S. Rajagopal, A. Barbeira, Y. Liang, O. Melia, L. Bastarache, Y. Park, The GTEx Consortium, X. Wen, H. K. Im, PhenomeXcan: Mapping the genome to the phenome through the transcriptome. *bioRxiv* (2019), p. 833210.

80. X. Wen, R. Pique-Regi, F. Luca, Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLoS Genet.* **13**, e1006646 (2017).

Acknowledgments:

We thank the donors and their families for their generous gifts of organ donation for transplantation, and tissue donations for the GTEx research project; the Genomics Platform at the Broad Institute for data generation; Jeffrey Struewing for his support and leadership of the GTEx project; Mariya Khan and Christopher Stolte for the illustrations in Figure 1; Bernhard HF Weber and Tobias Strunz for assistance with replicating the sb-eQTL for *HKDC1* in liver; Adam Gruenbaum for power estimation; Dan Nicolae and Lin Chen for advice on mediation analysis, Jacek Witkos for comments on an earlier version of the manuscript, and Mike Gloudemans for making the sex-stratified eQTL data available on LocusCompare (http://locuscompare.com/). This work was completed in part with computational resources provided by the Center for Research Informatics at The University of Chicago. The Center for Research Informatics is funded by the Biological Sciences Division at the University of Chicago with additional funding provided by the Institute for Translational Medicine, CTSA grant number UL1 TR000430 from the National Institutes of Health.

Author contributions: B.E.S. conceived the study; M.O. and B.E.S. led the writing and editing of the manuscript and supplement; M.O. and B.E.S. coordinated analyses of all contributing

authors; M.O. and V.W. performed differential gene expression analysis; B.B., D.J.C., M.M.-A., M.O., V.W. characterized effect sizes of sex-biased genes; M.M.-A., M.O., and V.W performed analysis of transcription factor binding sites; M.M.-A., M.O., and V.W performed tissue clustering based on gene expression levels and sex bias; M.M.-A. built the expression-based sex classifier; Y.Z. performed *MASH* analyses; M.S. and S.K.-H. provided advice on *MASH* analysis; M.O. generated sb-eQTL pipelines and performed sb-eQTL mapping; F.A., B.B., A.J.B., B.E.E., E.E., P.E., E.R.G., S.K.-H., S.K., E.A.K., S.B.M., P.P., A.D.S., and B.E.S. contributed to sb-eQTL analysis approach; D.J.C., S.K.-H., E.A.K., M.O., and V.W. characterized sb-eGenes; F.A., B.B. and A.V. performed sb-eQTL replication analysis in external datasets; S.K.-H., A.M.-P. and J.-M.S. contributed to sb-eQTL replication analysis; S.K. performed ASE aFC validation of sb-eQTLs; S.E.C., S.K.-H. and P.M. contributed to ASE aFC validation of sb-eQTLs; P.P. performed EAGLE ASE validation of sb-eQTLs; A.J.B. and S.K.-H. provided advice on EAGLE ASE validation; S.K.-H. performed coloc analysis; M.O. performed mediation analysis; S.K.-H. and B.L.P. provided advice on mediation analysis; F.A., D.G., S.K.-H., D.J.C., M.O., M.M.-A, and V.W. generated figures; F.A., K.G.A., and A.V.S. generated and oversaw GTEx v8 data generation, LDACC & pipelines; A.N.B., R.B., and H.K.I. generated GWAS data; F.A., S.K.-H., and M.O. generated cell type abundances and ieQTL data; S.K.-H., M.M.-A., M.O. characterized sex differences in cell type abundances; M.M.-A. and V.W. characterized phenotype relationships with cell type abundances; A.B., A.D.H.G., A.R.H., E.A.K., A.J.P., B.E.E, D.G., E.R.G., S.K.-H., A.M.-P., F.R., and A.D.S. performed analysis or provided feedback that significantly shaped this work but was not included in this final version; M.M.-A. and V.W. managed data and code in the GitHub repository; A.J.B., B.E.E., E.T.D., R.G., H.K.I., T.L., S.B.M., B.L.P., M.S., A.V.S., and B.E.S. supervised the work of trainees in his/her lab; D.J.C., S.K., S.K.-H., M.M.-A., M.O., P.P., V.W., Y.Z., and B.E.S. wrote manuscript text; B.B., A.J.B., B.E.E., A.D.H.G., R.G., S.K.-H., H.K.I., E.A.K., T.L., M.M.-A., M.O., S.B.M., L.C., S.K., P.P., B.L.P., A.D.S., B.E.S., A.V., and V.W., edited the manuscript. All authors read and approved the final manuscript.

GTEx Consortium*

Laboratory and Data Analysis Coordinating Center (LDACC): François Aguet[1], Shankara Anand[1], Kristin G Ardlie[1], Stacey Gabriel[1], Gad Getz[1,30], Aaron Graubert[1], Kane Hadley[1], Robert

E Handsaker[32,33,34], Katherine H Huang[1], Seva Kashin[32,33,34], Xiao Li[1], Daniel G MacArthur[33,35], Samuel R Meier[1], Jared L Nedzel[1], Duyen Y Nguyen[1], Ayellet V Segrè[1,17], Ellen Todres[1]

Analysis Working Group (funded by GTEx project grants): François Aguet[1], Shankara Anand[1], Kristin G Ardlie[1], Brunilda Balliu[40], Alvaro N Barbeira[2], Alexis Battle[18,11], Rodrigo Bonazzola[2], Andrew Brown[3,4], Christopher D Brown[24], Stephane E Castel[5,6], Don Conrad[41,42], Daniel J Cotter[29], Nancy Cox[16], Sayantan Das[26], Olivia M de Goede[29], Emmanouil T Dermitzakis[3,27,28], Barbara E Engelhardt[7,8], Eleazar Eskin[43], Tiffany Y Eulalio[44], Nicole M Ferraro[44], Elise Flynn[5,6], Laure Fresard[12], Eric R Gamazon[13,14,15,16], Diego Garrido-Martín[22], Nicole R Gay[29], Gad Getz[1,30], Aaron Graubert[1], Roderic Guigó[22,31], Kane Hadley[1], Andrew R Hamel[17,1], Robert E Handsaker[32,33,34], Yuan He[18], Paul J Hoffman[5], Farhad Hormozdiari1[9,1], Lei Hou[45,1], Katherine H Huang[1], Hae Kyung Im[2], Brian Jo[7,8], Silva Kasela[5,6], Seva Kashin[32,33,34], Manolis Kellis[45,1], Sarah Kim-Hellmuth[5,6,9], Alan Kwong[26], Tuuli Lappalainen[5,6], Xiao Li[1], Xin Li[12], Yanyu Liang[2], Daniel G MacArthur[33,35], Serghei Mangul[43,46], Samuel R Meier[1], Pejman Mohammadi[5,6,20,21,] Stephen B Montgomery[12,29], Manuel Muñoz-Aguirre[22,23], Daniel C Nachun[12], Jared L Nedzel[1], Duyen Y Nguyen[1], Andrew B Nobel[47], Meritxell Oliva[2,10], YoSon Park[24,25], Yongjin Park[45,1], Princy Parsana[11], Ferran Reverter[48], John M Rouhana[17,1], Chiara Sabatti[49], Ashis Saha[11], Ayellet V Segrè[1,17], Andrew D Skol[2,50], Matthew Stephens[36], Barbara E Stranger[2,37], Benjamin J Strober[18], Nicole A Teran[12], Ellen Todres[1], Ana Viñuela[38,3,27,28], Gao Wang[36], Xiaoquan Wen[26], Fred Wright[51], Valentin Wucher[22], Yuxin Zou[39]

Analysis Working Group (not funded by GTEx project grants): Pedro G Ferreira[52,53,54], Gen Li[55], Marta Melé[56], Esti Yeger-Lotem[57,58]

Leidos Biomedical - Project Management: Mary E Barcus[59], Debra Bradbury[60], Tanya Krubit[60], Jeffrey A McLean[60], Liqun Qi[60], Karna Robinson[60], Nancy V Roche[60], Anna M Smith[60], Leslie Sobin[60], David E Tabor[60], Anita Undale[60]

Biospecimen collection source sites: Jason Bridge[61], Lori E Brigham[62], Barbara A Foster[63], Bryan M Gillard[63], Richard Hasz[64], Marcus Hunter[65], Christopher Johns[66], Mark Johnson[67], Ellen Karasik[63], Gene Kopen[68], William F Leinweber[68], Alisa McDonald[68], Michael T Moser[63], Kevin

Myer[65], Kimberley D Ramsey[63], Brian Roe[65], Saboor Shad[68], Jeffrey A Thomas[68,67], Gary Walters[67], Michael Washington[67], Joseph Wheeler[66]

Biospecimen core resource: Scott D Jewell[69], Daniel C Rohrer[69], Dana R Valley[69]

Brain bank repository: David A Davis[70], Deborah C Mash[70] Pathology: Mary E Barcus[59], Philip A Branton[71], Leslie Sobin[60]

ELSI study: Laura K Barker[72], Heather M Gardiner[72], Maghboeba Mosavel[73], Laura A Siminoff[72]

Genome Browser Data Integration & Visualization: Paul Flicek[74], Maximilian Haeussler[75], Thomas Juettemann[74], W James Kent[75], Christopher M Lee[75], Conner C Powell[75], Kate R Rosenbloom[75], Magali Ruffier[74], Dan Sheppard[74], Kieron Taylor[74], Stephen J Trevanion[74], Daniel R Zerbino[74]

eGTEx groups: Nathan S Abell[29], Joshua Akey[76], Lin Chen[10], Kathryn Demanelis[10], Jennifer A Doherty[77], Andrew P Feinberg[78], Kasper D Hansen[79], Peter F Hickey[80], Lei Hou45,1, Farzana Jasmine[10], Lihua Jiang[29], Rajinder Kaul[81,82], Manolis Kellis[45,1], Muhammad G Kibriya[10], Jin Billy Li[29], Qin Li[29], Shin Lin[83], Sandra E Linder[29], Stephen B Montgomery[12,29], Meritxell Oliva[2,10], Yongjin Park[45,1], Brandon L Pierce[10], Lindsay F Rizzardi[84], Andrew D Skol[2,50], Kevin S Smith[12], Michael Snyder[29], John Stamatoyannopoulos[81,85], Barbara E Stranger[2,37], Hua Tang[29], Meng Wang[29]

NIH program management: Philip A Branton[71], Latarsha J Carithers[71,86], Ping Guan[71], Susan E Koester[87], A Roger Little[88], Helen M Moore[71], Concepcion R Nierras[89], Abhi K Rao[71], Jimmie B Vaught[71], Simona Volpi[90]

Affiliations

1. The Broad Institute of MIT and Harvard, Cambridge, MA, USA
2. Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL, USA
3. Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland
4. Population Health and Genomics, University of Dundee, Dundee, Scotland, UK

5. New York Genome Center, New York, NY, USA
6. Department of Systems Biology, Columbia University, New York, NY, USA
7. Department of Computer Science, Princeton University, Princeton, NJ, USA
8. Center for Statistics and Machine Learning, Princeton University, Princeton, NJ, USA
9. Statistical Genetics, Max Planck Institute of Psychiatry, Munich, Germany
10. Department of Public Health Sciences, The University of Chicago, Chicago, IL, USA
11. Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA
12. Department of Pathology, Stanford University, Stanford, CA, USA
13. Data Science Institute, Vanderbilt University, Nashville, TN, USA
14. Clare Hall, University of Cambridge, Cambridge, UK
15. MRC Epidemiology Unit, University of Cambridge, Cambridge, UK
16. Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
17. Ocular Genomics Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA
18. Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA
19. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
20. Scripps Research Translational Institute, La Jolla, CA, USA
21. Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA
22. Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Barcelona, Catalonia, Spain
23. Department of Statistics and Operations Research, Universitat Politècnica de Catalunya (UPC), Barcelona, Catalonia, Spain
24. Department of Genetics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA
25. Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA
26. Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA
27. Institute for Genetics and Genomics in Geneva (iGE3), University of Geneva, Geneva, Switzerland

28. Swiss Institute of Bioinformatics, Geneva, Switzerland
29. Department of Genetics, Stanford University, Stanford, CA, USA
30. Cancer Center and Department of Pathology, Massachusetts General Hospital, Boston, MA, USA
31. Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain
32. Department of Genetics, Harvard Medical School, Boston, MA, USA
33. Program in Medical and Population Genetics, The Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA
34. Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA
35. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA
36. Department of Human Genetics, University of Chicago, Chicago, IL, USA
37. Center for Genetic Medicine, Department of Pharmacology, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA
38. Department of Twin Research and Genetic Epidemiology, King's College London, London, UK
39. Department of Statistics, University of Chicago, Chicago, IL, USA
40. Department of Biomathematics, University of California, Los Angeles, Los Angeles, CA, USA
41. Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA
42. Department of Pathology & Immunology, Washington University School of Medicine, St. Louis, Missouri, USA
43. Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA
44. Program in Biomedical Informatics, Stanford University School of Medicine, Stanford, CA, USA
45. Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA
46. Department of Clinical Pharmacy, School of Pharmacy, University of Southern California, Los Angeles, CA, USA
47. Department of Statistics and Operations Research and Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA

48. Department of Genetics, Microbiology and Statistics, University of Barcelona, Barcelona. Spain.
49. Departments of Biomedical Data Science and Statistics, Stanford University, Stanford, CA, USA
50. Department of Pathology and Laboratory Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA
51. Bioinformatics Research Center and Departments of Statistics and Biological Sciences, North Carolina State University, Raleigh, NC, USA
52. Department of Computer Sciences, Faculty of Sciences, University of Porto, Porto, Portugal
53. Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal
54. Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal
55. Columbia University Mailman School of Public Health, New York, NY, USA
56. Life Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain
57. Department of Clinical Biochemistry and Pharmacology, Ben-Gurion University of the Negev, Beer-Sheva, Israel
58. National Institute for Biotechnology in the Negev, Beer-Sheva, Israel
59. Leidos Biomedical, Frederick, MD, USA
60. Leidos Biomedical, Rockville, MD, USA
61. UNYTS, Buffalo, NY, USA
62. Washington Regional Transplant Community, Annandale, VA, USA
63. Therapeutics, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA
64. Gift of Life Donor Program, Philadelphia, PA, USA
65. LifeGift, Houston, TX, USA
66. Center for Organ Recovery and Education, Pittsburgh, PA, USA
67. LifeNet Health, Virginia Beach, VA. USA
68. National Disease Research Interchange, Philadelphia, PA, USA
69. Van Andel Research Institute, Grand Rapids, MI, USA
70. Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA
71. Biorepositories and Biospecimen Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD, USA
72. Temple University, Philadelphia, PA, USA

73. Virgina Commonwealth University, Richmond, VA, USA
74. European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, United Kingdom
75. Genomics Institute, UC Santa Cruz, Santa Cruz, CA, USA
76. Carl Icahn Laboratory, Princeton University, Princeton, NJ, USA
77. Department of Population Health Sciences, The University of Utah, Salt Lake City, Utah, USA
78. Departments of Medicine, Biomedical Engineering, and Mental Health, Johns Hopkins University, Baltimore, MD, USA
79. Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA
80. Department of Medical Biology, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia
81. Altius Institute for Biomedical Sciences, Seattle, WA, USA
82. Division of Genetics, University of Washington, Seattle, WA, University of Washington, Seattle, WA, USA
83. Department of Cardiology, University of Washington, Seattle, WA, USA
84. HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA
85. Genome Sciences, University of Washington, Seattle, WA, USA
86. National Institute of Dental and Craniofacial Research, Bethesda, MD, USA
87. Division of Neuroscience and Basic Behavioral Science, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA
88. National Institute on Drug Abuse, Bethesda, MD, USA
89. Office of Strategic Coordination, Division of Program Coordination, Planning and Strategic Initiatives, Office of the Director, National Institutes of Health, Rockville, MD, USA
90. Division of Genomic Medicine, National Human Genome Research Institute, Bethesda, MD, USA

Funding:

The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health (NIH) and by the National Cancer Institute (NCI), the National

Human Genome Research Institute (NHGRI), the National Heart, Lung, and Blood Institute (NHLBI), the National Institute on Drug Abuse (NIDA), the National Institute of Mental Health (NIMH) and the National Institute of Neurological Disorders and Stroke (NINDS).

This work was funded by GTEx program grants: HHSN268201000029C (F.A., K.G.A., A.V.S., X.Li., E.T., S.G., A.G., S.A., K.H.H., D.Y.N., K.H., S.R.M., J.L.N.), 5U41HG009494 (F.A., K.G.A.), 10XS170 (Subcontract to Leidos Biomedical) (W.F.L., J.A.T., G.K., A.M., S.S., R.H., G.Wa., M.J., M.Wa., L.E.B., C.J., J.W., B.R., M.Hu., K.M., L.A.S., H.M.G., M.Mo., L.K.B.), 10XS171 (Subcontract to Leidos Biomedical) (B.A.F., M.T.M., E.K., B.M.G., K.D.R., J.B.), 10ST1035 (Subcontract to Leidos Biomedical) (S.D.J., D.C.R., D.R.V.), R01DA006227-17 (D.C.M., D.A.D.), Supplement to University of Miami grant DA006227. (D.C.M., D.A.D.), HHSN261200800001E (A.M.S., D.E.T., N.V.R., J.A.M., L.S., M.E.B., L.Q., T.K., D.B., K.R., A.U.), R01MH101814 (M.M-A., V.W., S.B.M., R.G., E.T.D., D.G-M., A.V.), U01HG007593 (S.B.M.), R01MH101822 (C.D.B.), U01HG007598 (M.O., A.D.S., B.E.S.), R01MH101820 (E.A.K., P.E., B.E.S), U01MH104393 (A.P.F.), as well as other funding sources: R01MH106842 (T.L., P.M., E.F., P.J.H.), R01HL142028 (T.L., Si.Ka., P.J.H.), R01GM122924 (T.L., S.E.C.), R01MH107666 (H.K.I.), P30DK020595 (H.K.I.), UM1HG008901 (T.L.), R01GM124486 (T.L.), R01HG010067 (Y.Pa.), R01HG002585 (G.Wa., M.St.), Gordon and Betty Moore Foundation GBMF 4559 (G.Wa., M.St.), 1K99HG009916-01 (S.E.C.), R01HG006855 (Se.Ka., R.E.H.), BIO2015-70777-P, Ministerio de Economia y Competitividad and FEDER funds (M.M-A., V.W., R.G., D.G-M.), NIH CTSA grant UL1TR002550-01 (P.M.), Marie-Skłodowska Curie fellowship H2020 Grant 706636 (S.K-H.), R35HG010718 (E.R.G.), FPU15/03635, Ministerio de Educación, Cultura y Deporte (M.M-A.), R01MH109905, 1R01HG010480 (A.Ba.), Searle Scholar Program (A.Ba.), R01HG008150 (S.B.M.), 5T32HG000044-22, NHGRI Institutional Training Grant in Genome Science (N.R.G.), EU IMI program (UE7-DIRECT-115317-1) (E.T.D., A.V.), FNS funded project RNA1 (31003A_149984) (E.T.D., A.V.), DK110919 (F.H.), F32HG009987 (F.H.), Consolidate Research Group. Generalitat de Catalunya SGR 1736 and CERCA program (A.M.-P., J.M.S.); Rhodes Trust and Natural Sciences and Engineering Research Council of Canada (A.J.P.). All CRG authors acknowledge the support of the Spanish Ministry of Science, Innovation and Universities to the EMBL partnership, the Centro de Excelencia Severo Ochoa and the CERCA Programme / Generalitat de Catalunya.

Competing interests: F.A. is an inventor on a patent application related to TensorQTL; S.E.C. is a co-founder, chief technology officer and stock owner at Variant Bio; D.G.M. is co-founder with equity in Goldfinch Bio, and has received research support from AbbVie, Astellas, Biogen, BioMarin, Eisai, Merck, Pfizer, and Sanofi-Genzyme; E.A.K. is an employee of Janssen Pharmaceuticals; H.I. has received speaker honoraria from GSK and AbbVie; E.T.D. is chairman and member of the board of Hybridstat LTD; G.G. receives research funds from IBM and Pharmacyclics, and is an inventor on patent applications related to MuTect, ABSOLUTE, MutSig, POLYSOLVER and TensorQTL; T.L. is a scientific advisory board member of Variant Bio with equity, and Goldfinch Bio.

Data and materials availability: All GTEx open-access data, including summary statistics and visualizations of sex-biased eQTLs, will be released on the GTEx Portal (https://gtexportal.org/home/datasets). GTEx v8 sex-stratified eQTL data will be available on LocusCompare (http://locuscompare.com/). All GTEx protected data are available via dbGaP (accession phs000424.v8). Access to the raw sequence data is now provided through the AnVIL platform (https://gtexportal.org/home/protectedDataAccess). The QTL mapping pipeline is available at https://github.com/broadinstitute/gtex-pipeline. Residual GTEx biospecimens have been banked, and remain available as a resource for further studies (access can be requested on the GTEx Portal, at https://www.gtexportal.org/home/samplesPage).

Supplementary Materials:

Materials and Methods

Figures S1-S9

Tables S1-S15

References (*58-80*)